

GenCore version 5.1.3
Copyright (c) 1993 : 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2002, 15:04:13 : Search time 4357 Seconds
(without alignments)
11341.878 Million cell updates/sec

Title: HSYBL_COPY_453_2150
Perfect score: 1698
Sequence: 1 AGAACCATATGTCGGCCAA.....CACCCCGGAGGAGCGCCAG 1698

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**
2: gb_hgt:**
3: gb_in:**
4: gb_om:**
5: gb_ov:**
6: gb_pat:**
7: gb_ph:**
8: gb_pl:**
9: gb_pr:**
10: gb_ro:**
11: gb_sts:**
12: gb_sy:**
13: gb_un:**
14: gb_vi:**
15: em_ba:**
16: em_fun:**
17: em_hum:**
18: em_in:**
19: em_mu:**
20: em_om:**
21: em_or:**
22: em_ov:**
23: em_pat:**
24: em_ph:**
25: em_pl:**
26: em_ro:**
27: em_sts:**
28: em_un:**
29: em_vi:**
30: em_hgt_hum:**
31: em_hgt_inv:**
32: em_hgt_other:**
33: em_hgt_mus:**
34: em_hgt_pln:**
35: em_hgt_rdd:**
36: em_hgt_man:**
37: em_hgt_vrt:**
38: em_sy:**
39: em_hgtgo_hum:**
40: em_hgtgo_mus:**
41: em_hgtgo_other:**

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	1698	100.0	2550	9	HSVB1	X95666 H.sapiens Y	
2	1672.4	98.5	161835	9	AC098484	AC098484 Homo sapi	
3	456.6	26.9	659	9	HUMY232E04	AF086102 Homo sapi	
C	4	186.8	11.0	195623	2	AC101988	AC101988 Mus muscu
5	107.4	6.3	1548	6	AX014868	AX014868 Sequence	
6	101.4	6.0	144902	2	AC016114	AC016114 Homo sapi	
7	101.4	6.0	162144	9	CNS06C85	AL391261 Human chr	
C	8	101.4	6.0	184656	9	CNS05TDB	AL356021 Human chr
C	9	96.6	5.7	189367	9	AL162733	AL162733 Human DNA
10	95	5.6	99886	9	AL135841	AL135841 Human DNA	
11	89.4	5.3	395	6	AX321258	AX321258 Sequence	
12	89	5.2	1481	6	AR117695	AR117695 Sequence	
13	89	5.2	1481	6	AX409453	AX409453 Sequence	
14	89	5.2	1481	9	HUMYB1A	J03827 Y box bindi	
15	87.4	5.1	2168	9	HUMPSDBPB	L37516 Homo sapien	
16	86.6	5.1	162996	2	AC051655	AC051655 Homo sapi	
17	85.2	5.0	410	6	AX321869	AX321869 Sequence	
18	85.2	5.0	926	9	BC000064	BC000064 Homo sapi	
19	83	4.9	3073	6	AR083654	AR083654 Sequence	
20	83	4.9	3073	6	I22491	I22491 Sequence 31	
21	81.4	4.8	1468	9	HUMNSEP	M83234 Human nucle	
22	81.4	4.8	1474	9	HUMNRABP	M85234 Human nucle	
23	80.6	4.7	142184	2	AC037428	AC037428 Homo sapi	
24	80.6	4.7	143604	9	AC093610	AC093610 Homo sapi	
25	80.6	4.7	310021	2	AC092883	AC092883 Homo sapi	
26	79	4.7	1552	9	BC002411	BC002411 Homo sapi	
C	27	79	4.7	77948	2	AC018982	AC018982 Homo sapi
28	78.8	4.6	381	6	AX365822	AX365822 Sequence	
29	77.8	4.6	241390	2	AC093965	AC093965 Rattus no	
30	77.4	4.6	545	6	AX321285	AX321285 Sequence	
31	77	4.5	1554	9	BC010430	BC010430 Homo sapi	
32	76.8	4.5	142565	10	AL592545	AL592545 Mouse DNA	
33	76.8	4.5	197658	2	AL646094	AL646094 Mus muscu	
34	76.6	4.5	7218	6	I66494	I66494 Sequence 14	
35	76.6	4.5	179402	9	AC079328	AC079328 Homo sapi	
C	36	76.6	4.5	187005	2	AC024309	AC024309 Homo sapi
37	75.4	4.4	1452	9	HUMDBPB	M24070 Human DNA-b	
38	74	4.4	307	9	HSBFB31	Z70765 H.sapiens m	
39	73.8	4.3	1503	4	OCU16821	OCU16821 Oryctolagus	
C	40	72.8	4.3	169224	2	AC113562	AC113562 Sus scrofa
C	41	72.6	4.3	159409	2	AC092900	AC092900 Homo sapi
42	72.6	4.3	160877	2	AC026560	AC026560 Homo sapi	
C	43	72.6	4.3	186333	2	AC027493	AC027493 Homo sapi
C	44	72.6	4.3	187215	9	AC092915	AC092915 Homo sapi
45	67.4	4.0	1407	10	BC013450	BC013450 Mus muscu	

ALIGNMENTS

RESULT 1

HSYB1
LOCUS HSYB1 2550 bp DNA linear PRI 08-JUL-1996
DEFINITION H. sapiens YB-1 gene promoter region.
ACCESSION X96666
VERSION X96666.1 GI:1403348
KEYWORDS promoter region; Y box binding protein; YB-1 gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2550)
AUTHORS Makino, Y., Ohga, T., Toh, S., Koike, K., Okumura, K., Wada, M.,
Kuwano, M. and Kohno, K.
TITLE Structural and functional analysis of the human Y-box binding

protein (YB-1) gene promoter
Nucleic Acids Res. 24 (10), 1873-1878 (1996)
96226173
PUBMED
8657568
REFERENCE
2 (bases 1 to 2550)
Yoshinari M.
Direct Submission
Submitted (15-MAR-1996) Yoshinari M., Department of Biochemistry,
Kyushu University School of Medicine, Maidashi, Fukuoka, 812-82,
JAPAN
JOURNAL

FEATURES
source
1..2550
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2411
/gene="YB-1"
1..1855
/gene="YB-1"
1856..2352
/gene="YB-1"
/number=1
1904..1914
/rpt_type=INVERTED
1957..1962
/gene="YB-1"
/bound_moiety="SP-1"
2001..2006
/gene="YB-1"
2024..2029
/bound_moiety="SP-1"
/gene="YB-1"
2033..2038
/bound_moiety="SP-1"
/gene="YB-1"
2034..2044
/rpt_type=INVERTED
2100..2105
/gene="YB-1"
/bound_moiety="SP-1"
2187..2411
/gene="YB-1"
/codon_start=1
/protein_id="CAA65446.1"
/db_xref="GI:1403349"
/db_xref="SPTREMBL:Q15905"
/translation="MSSEAEIQPPAAPALSAADTKPQTGGAGSGPGGLTS
AAPAGDKKIVIGDRQGRGWGPRAAQQRNR"
BASE COUNT 611 a 708 c 660 g 571 t
ORIGIN

Query Match 100.0%; Score 1698; DB 9; Length 2550;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAACCATATGGTCGCCAGCTATATTATTAACCTATTGGCTCTCACAGAAAAAGGTT 60
|||||
Db 453 AGAACCATATGGTCGCCAGCTATATTATTAACCTATTGGCTCTCACAGAAAAAGGTT 512
QY 61 CCTACCCCTTGCTCTAAGCAATTGGTGATGATGGCTACAGTAGTGACACTAGGAATGAA 120
|||||
Db 513 CCTACCCCTTGCTCTAAGCAATTGGTGATGATGGCTACAGTAGTGACACTAGGAATGAA 572
QY 121 AAGACTCTGAATTGACATTTAGAGGTTTAGGACTGACTCAAGAAGCCCTTTAGGAGGTG 180
|||||
Db 573 AAGACTCTGAATTGACATTTAGAGGTTTAGGACTGACTCAAGAAGCCCTTTAGGAGGTG 632
QY 181 GAACCTACAGGCTACAGCGCATTTAGGGTTAGGGTTAGGAATAGTATAGCTCTCTAACT 240
|||||
Db 633 GAACCTACAGGCTACAGCGCATTTAGGGTTAGGGTTAGGAATAGTATAGCTCTCTAACT 692
QY 241 GCCTTCTCTACTTCCATTCCTGCCCTCTGCAACCCATTCTCCACTCCGCGCCATTTT 300
|||||

Db 693 GCCTTCTCTACTTCCATTCCTTGGCCCTCTGCAAOCCATTCTCCACTCCGAGCCATTTT 752
QY 301 TAAAAAGATGCCCTCCCTACTTACTGACTCTAAAATTGTCTTCTCTACTCTCCCTCAG 360
|||||
Db 753 TAAAAAGATGCCCTCCCTACTTACTGACTCTAAAATTGTCTTCTCTACTCTCCCTCAG 812
QY 361 GATATATTTTCCAAATTAATATACCTAAGTAGCTGCGCCACTCTGCAACCAATGTGCAT 420
|||||
Db 813 GATATATTTTCCAAATTAATATACCTAAGTAGCTGCGCCACTCTGCAACCAATGTGCAT 872
QY 421 TCGAGTCTTACTGAACACTTGTGACTGCATTTCCCGAGATCTCACCTCTTCTCGCCTGTAC 480
|||||
Db 873 TCGAGTCTTACTGAACACTTGTGACTGCATTTCCCGAGATCTCACCTCTTCTCGCCTGTAC 932
QY 481 CTTGTGCGGGAAGTACGCCCTCCACCTTCTCCCTGTCTTCCACTCCCAAAATACTTCTGT 540
Db 933 CTTGTGCGGGAAGTACGCCCTCCACCTTCTCCCTGTCTTCCACTCCCAAAATACTTCTGT 992
QY 541 GTTTTTGCAGCTCTGAGTAGTATTTACCGTGTGGCTGTTTAAATTTCTGCTCTCATCAGAA 600
|||||
Db 993 GTTTTTGCAGCTCTGAGTAGTATTTACCGTGTGGCTGTTTAAATTTCTGCTCTCATCAGAA 1052
QY 601 GGCAGAAACTGACTCGGGAAGTATTTCCATCCCGAGCGGATAGTAGCGCTTAAAAAGAA 660
|||||
Db 1053 GGCAGAAACTGACTCGGGAAGTATTTCCATCCCGAGCGGATAGTAGCGCTTAAAAAGAA 1112
QY 661 CGGAAGAAGTGGTGGGAGGACTTTCAGTAACATCAGGTGGAGCGCTCAATTTTATTCGTT 720
Db 1113 CGGAAGAAGTGGTGGGAGGACTTTCAGTAACATCAGGTGGAGCGCTCAATTTTATTCGTT 1172
QY 721 TGTGAACGTGGATAGTAATCCCTCTATCACGTGGCTGTGGCAGGAATAAAGTGAAGAAA 780
|||||
Db 1173 TGTGAACGTGGATAGTAATCCCTCTATCACGTGGCTGTGGCAGGAATAAAGTGAAGAAA 1232
QY 781 CAAACAGGCTAGCTTGTTCATAAATGTAGTTGAATTAATCTGATTTGTGGTCAGTA 840
Db 1233 CAAACAGGCTAGCTTGTTCATAAATGTAGTTGAATTAATCTGATTTGTGGTCAGTA 1292
QY 841 GAAAAAGATGTGAATACTTGGAAAGGAGACACATTTTTTAAATATATGCCTTGTGTA 900
Db 1293 GAAAAAGATGTGAATACTTGGAAAGGAGACACATTTTTTAAATATATGCCTTGTGTA 1352
QY 901 CGGATCAGAAGCAGTCCCCCATGGAGCACACCTCGCCCTTAACATGCTGAACCCGGC 960
Db 1353 CGGATCAGAAGCAGTCCCCCATGGAGCACACCTCGCCCTTAACATGCTGAACCCGGC 1412
QY 961 TGCATAGGCTCGGTGGTCCCTCAAGGTGACTGTCTCCGACAAAAGGGTACGCTCTTCAA 1020
Db 1413 TGCATAGGCTCGGTGGTCCCTCAAGGTGACTGTCTCCGACAAAAGGGTACGCTCTTCAA 1472
QY 1021 ACGCATACGTTTAAAGCAATTCAGAAACCCTCGGCTGTGCCGCGACTACACGGCCATTA 1080
Db 1473 ACGCATACGTTTAAAGCAATTCAGAAACCCTCGGCTGTGCCGCGACTACACGGCCATTA 1532
QY 1081 AGAAAAGAGGACTCTATGCCCGCCGTAATGTCTCAGATCAGAGGAGCGTATTTGGAG 1140
Db 1533 AGAAAAGAGGACTCTATGCCCGCCGTAATGTCTCAGATCAGAGGAGCGTATTTGGAG 1592
QY 1141 CTGGGAGGAGGGAAGCCCTTTTCTCACGGGGGTAAAGCGCTTTCGAGCCCTTCCA 1200
Db 1593 CTGGGAGGAGGGAAGCCCTTTTCTCACGGGGGTAAAGCGCTTTCGAGCCCTTCCA 1652
QY 1201 ATCCCGGGTCCGCGCGGTAAATCCCTCGCCAGCGTTCGGGCGGTGCCTTTTTTCAGCCGA 1260
Db 1653 ATCCCGGGTCCGCGCGGTAAATCCCTCGCCAGCGTTCGGGCGGTGCCTTTTTTCAGCCGA 1712
QY 1261 GACACAACTTGAACGTGGGGGCGCCAGCGCGGTGCCTCTGTGGAAGTCAAGTTC 1320
Db 1713 GACACAACTTGAACGTGGGGGCGCCAGCGCGGTGCCTCTGTGGAAGTCAAGTTC 1772
QY 1321 CTCTCTCGCTCTCTCGGTACTCTATGTTTTCGTGGCGGACTACTCTAATCTAGTT 1380
Db 1773 CTCTCTCGCTCTCTCGGTACTCTATGTTTTCGTGGCGGACTACTCTAATCTAGTT 1832

3398	3599	3657	3675	11517	11316	BASE COUNT	43236	a	36046	c	37167	g	45386	t	ORIGIN
2503	2527	14158	13689	1044	1024	Query Match	98.5%	Score	1672.4	DB	9	Length	161835		
3341	3393	2386	2336	5448	5071	Best Local Similarity	99.8%	Pred. No.	0	Mismatches	1	Indels	2	Gaps	2
5997	5970	4939	5026	2779	2792	Matches 1695	Conservative	0							
11401	11247	2187	2206	3013	3120	QY	1	AGAACCATATGGTCGCGCAAGCTTATATTATTTGGCTCTCACAGAAAAAGTTT	60						
1470	1443	6457	6489	3869	3810	Db	28812	AGAACCATATGGTCGCGCAAGCTTATATTATTTGGCTCTCACAGAAAAAGTTT	28871						
12564	12727	1771	1836	9708	9870	QY	61	CCTACCCCTTGCCTCTAAGCAATTTGGTGATGATGGCTTACAGTAGTACAGTAGGAAATGAA	120						
6660	6745	3788	3991	12923	12794	Db	28872	CCTACCCCTTGCCTCTAAGCAATTTGGTGATGATGGCTTACAGTAGTACAGTAGGAAATGAA	28931						
4555	4619	4838	4792	4563	4622	QY	121	AAGACTCTGAATTCACATTTAGAGGGTTTAGGACTGACTCAAGAACCCCTTTAGGAGGTG	180						
1729	1734	3160	3127	287	<800	Db	28932	AAGACTCTGAATTCACATTTAGAGGGTTTAGGACTGACTCAAGAACCCCTTTAGGAGGTG	28991						
406	<800	4088	4313	9519	9493	QY	181	GAACCTCACAGCCCTAGACGGCATTGGTAGGGTAAGCAATTCGACTTCCGACGAGCCCTAACT	240						
2213	2193	5832	6201	2229	2229	Db	28992	GAACCTCACAGCCCTAGACGGCATTGGTAGGGTAAGCAATTCGACTTCCGACGAGCCCTAACT	29051						
6922	6981	942	971	2610	2628	QY	241	GCCTTCTCTACTTCCATTCCTGCCCTCTGCAACCCATTCTCCACTTCCGACGAGCCCTAACT	300						
4531	4619	4341	4619	5661	5695	Db	29052	GCCTTCTCTACTTCCATTCCTGCCCTCTGCAACCCATTCTCCACTTCCGACGAGCCCTAACT	29111						
1381	1370	9735	9702	576	<800	QY	301	TAAAAAGATGCCCTCCCTACTTATGACTTAAATTTGCTTCTCACTTCCGACGAGCCCTAACT	360						
1386	1370	4886	4792	820	830	Db	29112	TAAAAAGATGCCCTCCCTACTTATGACTTAAATTTGCTTCTCACTTCCGACGAGCCCTAACT	29171						
1168	1168	15773	15940	3668	3654	QY	361	GATATATTTCCCAATTAATATACCTAAGTACCTGACCTGCAACCCATTCTCCACTTCCGACGAGCCCTAACT	420						
15253	15479	398	<800	9025	8972	Db	29172	GATATATTTCCCAATTAATATACCTAAGTACCTGACCTGCAACCCATTCTCCACTTCCGACGAGCCCTAACT	29231						
3613	3599	5122	5509	5092	5071	QY	421	TCGAGTCTTACTGAATCTTACTGACTTCCGAGATCTCACTTCTCCGCTGCTAC	480						
8174	8090	1336	1324	1026	1024	Db	29232	TCGAGTCTTACTGAATCTTACTGACTTCCGAGATCTCACTTCTCCGCTGCTAC	29291						
2074	2086	10995	10821			QY	481	CCTGTGCGGGAAGTCAGCCCTCCGCTTCCCTGCTTCCACTCCCAAAATACCTTCGT	540						
714	<800	4329	4313			Db	29292	CCTGTGCGGGAAGTCAGCCCTCCGCTTCCCTGCTTCCACTCCCAAAATACCTTCGT	29351						
6275	6239	16	<800			QY	541	GGTTTTCAGCTCTGGAGTATTTACCGTGTGGCTGTTTAAATTTCTGCTTCCACTCCCAAAATACCTTCGT	600						
3627	3599	1338	1324			Db	29352	GGTTTTCAGCTCTGGAGTATTTACCGTGTGGCTGTTTAAATTTCTGCTTCCACTCCCAAAATACCTTCGT	29411						
		978	971			QY	601	GGCAGAACTGACTCGGAACTATTCCATCCCGAGCGGATAGTAGACGCTTAAAAAAGAA	660						
		2203	2206			Db	29412	GGCAGAACTGACTCGGAACTATTCCATCCCGAGCGGATAGTAGACGCTTAAAAAAGAA	29471						
		2343	2336			QY	661	CGAAGAAGTGGTGGGAGGACTTCAGTAACATCAGGTGGCAGCCCTCAATTTATCGTT	720						
		489	<800			Db	29472	CGAAGAAGTGGTGGGAGGACTTCAGTAACATCAGGTGGCAGCCCTCAATTTATCGTT	29530						
		6300	6489			QY	721	TCTGAAACGTGGATAGTAACTCCCTCTATCACGTGGTGTTCAGGAATAAAGTGAAGAAA	780						
		376	<800			Db	29531	TCTGAAACGTGGATAGTAACTCCCTCTATCACGTGGTGTTCAGGAATAAAGTGAAGAAA	29590						
		2981	2969			QY	781	CAAAACAGGCTAGCTTGTTCATAATAATGTGAGTTGAATTTAAATCTGAATTTGGTGCAGTA	840						
						Db	29591	CAAAACAGGCTAGCTTGTTCATAATAATGTGAGTTGAATTTAAATCTGAATTTGGTGCAGTA	29650						
						QY	841	GAAGAAGATGCAATACCTGGAAGGAAGACACACATTTTTTAAATATATGCTCGTGGTAA	900						
						Db	29651	GAAGAAGATGCAATACCTGGAAGGAAGACACACATTTTTTAAATATATGCTCGTGGTAA	29710						
						QY	901	CGGATCAGAAGGAGGTCCTCCATGGAGCACACCTCGCCCTAAACATGCTGAACCCGGGC	960						
						Db	29711	CGGATCAGAAGGAGGTCCTCCATGGAGCACACCTCGCCCTAAACATGCTGAACCCGGGC	29770						
						QY	961	TCCATAGCCTCGGTGGTCCCTCCCAAGGTGACTGCTCCGACAAAAAGGTCACCTCTCAA	1020						

FEATURES
source

Location/Qualifiers

1. .161835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP5-994D16"
/clone_lib="RPCI human PAC library 5"
1. .4
/note="single subclone region"
161830. .161835
/note="single subclone region"

misc_feature

misc_feature

Db 29771 TGCCATAGCTCGTGGTCCCTCCCAAGGTACTGTCTCCGACAAAGGGTACGCTCTTCAA 29830
QY 1021 ACGCATACGTTTAAAGCAATTCAGAAACCTTCGCGCTGTGCGCGACTACACGGCCATTA 1080
Db 29831 ACGCATACGTTTAAAGCAATTCAGAAACCTTCGCGCTGTGCGCGACTACACGGCCATTA 29890
QY 1081 AAGAAAGACGACTATGCCCCCGGTAATGTTCTCAGATACACGAGCCGTAATTTGGAG 1140
Db 29891 AAGAAAGACGACTATGCCCCCGGTAATGTTCTCAGATACACGAGCCGTAATTTGGAG 29950
QY 1141 CTGGAGGAGGAGGAGCCCTTCTTTCACGGGGGCTTAAGCGCTTCCTGAGCCCTTCCA 1200
Db 29951 CTGGAGGAGGAGGAGCCCTTCTTTCACGGGGGCTTAAGCGCTTCCTGAGCCCTTCCA 30010
QY 1201 ATCCCGGTCCGCGCGGTAAATCCCTGCCAGCGTTCGGCGTCCCTTTTTCAGCCGA 1260
Db 30011 ATCCCGGTCCGCGCGGTAAATCCCTGCCAGCGTTCGGCGTCCCTTTTTCAGCCGA 30070
QY 1261 GACACAACCTGAACGTGGGGGCGGCCAGCCCGCGGTCTCTGGAAGTCACGTTTC 1320
Db 30071 GACACAACCTGAACGT-GGGGGCGGCCAGCCCGCGGTCTCTGGAAGTCACGTTTC 30129
QY 1321 CTCTCGCGCTCTCTCGGCTACTCTATGTTTCTGTCGCCGACTACTCTAATTTAGTT 1380
Db 30130 CTCTCGCGCTCTCTCGGCTACTCTATGTTTCTGTCGCCGACTACTCTAATTTAGTT 30189
QY 1381 CCGGTCTCTATGCGCGCGCGGAGGAGCAGACCGTTGTAGTGCAGTCAATTTAGCGGCC 1440
Db 30190 CCGGTCTCTATGCGCGCGCGGAGGAGCAGACCGTTGTAGTGCAGTCAATTTAGCGGCC 30249
QY 1441 AAGGTTCCATGAGATGAGGAGTATGATTAATATTAGCCCATAGAGCTAGGATTTGG 1500
Db 30250 AAGGTTCCATGAGATGAGGAGTATGATTAATATTAGCCCATAGAGCTAGGATTTGG 30309
QY 1501 GTGAGTGGGACGATTCACAGTACCACTGCGCAGTGAACACCGCTTAGCGGGTCTGCTC 1560
Db 30310 GTGAGTGGGACGATTCACAGTACCACTGCGCAGTGAACACCGCTTAGCGGGTCTGCTC 30369
QY 1561 GTAGGCTTATCCCGCTGTGCGCGCATTCCTCGCTAGTTCGATCGGTAGCGGAGCGGAG 1620
Db 30370 GTAGGCTTATCCCGCTGTGCGCGCATTCCTCGCTAGTTCGATCGGTAGCGGAGCGGAG 30429
QY 1621 AGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGCGCTAGTTACCATCACA 1680
Db 30430 AGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGCGCTAGTTACCATCACA 30489
QY 1681 CCCCGGAGGAGCGCGAG 1698
Db 30490 CCCCGGAGGAGCGCGAG 30507

RESULT 3
HUMY292E04
LOCUS HUMY292E04 659 bp mRNA linear PRI 29-AUG-1998
DEFINITION Homo sapiens full length insert cDNA clone YZ92E04.
ACCESSION AF086102
VERSION AF086102.1 GI:3483447
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 659)
Woesner, J., Tan, F., Marra, M., Kucaba, T., Yandell, M., Martin, J.,
Marth, G., Bowles, L., Wylie, T., Bowers, Y., Steptoe, M., Theising, B.,
Geisel, S., Allen, M., Underwood, K., Chappell, J., Person, B.,
Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,
Schurk, R., Ritter, E., Kohn, S., Swaller, T., Behnmer, K., Hillier, L.,
Wilson, R., and Waterston, R.
Full Clone Sequencing of the Longest Available Member from Each
Unigene Cluster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 659)

AUTHORS
TITLE
JOURNAL

Waterston, R.
Direct Submission
Submitted (24-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@watson.wustl.edu

COMMENT

NOTICE: This sequence represents the full insert of this cDNA. No
attempt has been made to verify whether this corresponds to the
full-length of the original mRNA from which it was derived. We
have tried to obtain double-stranded, or double chemistry sequence
across the entire clone, but potentially, there are areas in the
sequence where this level of coverage was not achieved.
Nevertheless, we are confident of the accuracy of this sequence as
all regions of low quality, as defined by PHRAP (P. Green, in
preparation), were visually inspected and edited accordingly. The
consensus quality values for this sequence have been submitted
separately.

The location of this clone is unknown.

FEATURES
source

Location/Qualifiers
1..659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:290526"
/clone_lib="Soares_multiple_sclerosis_2NbhMSP"
repeat_region 20..119
/rpt_family="L2"
BASE COUNT 194 a 145 c 138 g 182 t
ORIGIN

Query Match 26.9%; Score 456.6; DB 9; Length 659;
Best Local Similarity 98.9%; Pred. No. 6.6e-126;
Matches 470; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 353 CCCTCAGGATATTTCCAAATTAATATACCTAAGTGAAGTGGCCACCTCTGCAACCCAA 412
Db 177 CACTTGAGGATATTTCCAAATTAATATACCTAAGTGAAGTGGCCACCTCTGCAACCCAA 236
QY 413 TGTCACATTCGAGTCTTACTGAACACTACTGACATTCATTCCTCCGAGATCTCACCTCTCTC 472
Db 237 TGTCACATTCGAGTCTTACTGAACACTACTGACATTCATTCCTCCGAGATCTCACCTCTCTC 296
QY 473 GCCTGTACCTGTGCGCGGAAAGTCAGCCCTCCACCTCTCTCCCTGCTTCCACCTCCCAAAA 532
Db 297 GCCTGTACCTGTGCGCGGAAAGTCAGCCCTCCACCTCTCTCCCTGCTTCCACCTCCCAAAA 356
QY 533 TACTTCGTGCTTTTGAGCTCTGAGTATTTACCGTCTTGGCTGTTAAATTTCTGCCTC 592
Db 357 TACTTCGTGCTTTTGAGCTCTGAGTATTTACCGTCTTGGCTGTTAAATTTCTGCCTC 416
QY 593 CATCAGAAGCGAGAAACTGACTCCGGAATATTTCCATCCCGAGCGGATAGTAGCGCTTA 652
Db 417 CATCAGAAGCGAGAAACTGACTCCGGAATATTTCCATCCCGAGCGGATAGTAGCGCTTA 476
QY 653 AAAAAGAACGAGAGGTTGGTGGGAGGACTTCAGTAACATCAGGTGGGAGCTCAATT 712
Db 477 AAAAAGAACGAGAGGTTGGTGGGAGGACTTCAGTAACATCAGGTGGGAGCTCAATT 535
QY 713 TTATCGTTTCTGAACGCTGATAGTATCTCTATCACCTGCTGTTGCGAGGATTAAG 772
Db 536 TTATCGTTTCTGAACGCTGATAGTATCTCTATCACCTGCTGTTGCGAGGATTAAG 595
QY 773 TGAATAACAAACAGGCTAGCTTGTTCATTAATGTGAGTTGAATTAATCTGA 827
Db 596 TGAATAACAAACAGGCTAGCTTGTTCATTAATGTGAGTTGAATTAATCTGA 650

RESULT 4

AC101988/c

LOCUS
DEFINITION

AC101988 195623 bp DNA linear HTG 21-AUG-2002
Mus musculus clone RP24-363D14, WORKING DRAFT SEQUENCE, 30
unordered pieces.

AC101988
AC101988.2

GI:22381361

VERSION
KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE
ORGANISM

house mouse.

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE
JOURNAL

1 (bases 1 to 195623)
Birren,B., Nusbaum,C. and Lander,E.

REFERENCE
AUTHORS

2 (bases 1 to 195623)
Unpublished

REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepe,I., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 195623)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,I., Collamore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17060764.

COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L17831

Center Clone name: 363_D14

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 185798 bases at least Q40

Consensus quality: 189816 bases at least Q30

Consensus quality: 191524 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 192723; sum-of-contigs

Quality coverage: 6.6 in Q20 bases; agarose-fp

Quality coverage: 6.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 637: contig of 637 bp in length
* 638 737: gap of 100 bp
* 738 1300: contig of 563 bp in length
* 1301 1400: gap of 100 bp
* 1401 2264: contig of 864 bp in length
* 2265 2364: gap of 100 bp
* 2365 3575: contig of 1211 bp in length
* 3576 3675: gap of 100 bp
* 3676 4828: contig of 1153 bp in length
* 4829 4928: gap of 100 bp
* 4929 5428: contig of 500 bp in length
* 5429 5528: gap of 100 bp
* 5529 6449: contig of 921 bp in length
* 6450 6549: gap of 100 bp
* 6550 7795: contig of 1246 bp in length
* 7796 7895: gap of 100 bp
* 7896 8117: contig of 222 bp in length
* 8118 8217: gap of 100 bp
* 8218 9112: contig of 895 bp in length
* 9113 9212: gap of 100 bp
* 9213 10260: contig of 1048 bp in length
* 10261 10360: gap of 100 bp
* 10361 13254: contig of 2894 bp in length
* 13255 13354: gap of 100 bp
* 13355 16381: contig of 3027 bp in length
* 16382 16481: gap of 100 bp
* 16482 18270: contig of 1789 bp in length
* 18271 18370: gap of 100 bp
* 18371 21470: contig of 3100 bp in length
* 21471 21570: gap of 100 bp
* 21571 22758: contig of 1188 bp in length
* 22759 22858: gap of 100 bp
* 22859 26779: contig of 3821 bp in length
* 26780 26779: gap of 100 bp
* 26780 60602: contig of 33823 bp in length
* 60603 60702: gap of 100 bp
* 60703 65030: contig of 4328 bp in length
* 65031 65130: gap of 100 bp
* 65131 69004: contig of 3874 bp in length
* 69005 69104: gap of 100 bp
* 69105 74370: contig of 5266 bp in length
* 74371 74470: gap of 100 bp
* 74471 82359: contig of 7889 bp in length
* 82360 82459: gap of 100 bp
* 82460 91758: contig of 9299 bp in length
* 91759 91858: gap of 100 bp
* 91859 100274: contig of 8416 bp in length
* 100275 100374: gap of 100 bp
* 100375 113878: contig of 13504 bp in length
* 113879 113978: gap of 100 bp

RESULT 10	Al135841	99886 bp	DNA	linear	PRI 05-JUL-2000
LOCUS	Al135841/c				
DEFINITION	Human DNA sequence from clone RP11-327L3 on chromosome 9p13.1-13.3. Contains a putative novel gene, a PGAM1 (phosphoglycerate mutase 1 (brain)) pseudogene, the gene for a novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein, an NSEP1 (nuclease sensitive element binding protein 1) (YB1) pseudogene, ESTs, STSS and GSSs, complete sequence.				
ACCESSION	Al135841				
VERSION	Al135841.11	GI:7529194			
KEYWORDS	HTG: NSEP1; olfactory receptor; PGAM1; rhodopsin; YB1.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 99886)				
TITLE	Corby, N.				
JOURNAL	Direct Submission				
COMMENT	Submitted (20-APR-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone Requests: clonequests@sanger.ac.uk On Aug 9, 2000 this sequence version replaced gi:7378508. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at:				

on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

RP11-327L3 is from the library: RP01-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-327L3. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-327L3 is at 99886 in this sequence. The true left end of clone RP11-113A10 is at 64692 in this sequence. The true right end of clone RP11-112J3 is at 100 in this sequence.

Location/Qualifiers

1. 99886
forward-biased-clone

FEATURES

Source

```

1. 99886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="p13.1-13.3"
/clone="RP11-327L3"
/clone_lib="RPCI-11.2"
4. 752
/note="L1M4c repeat: matches 1256. .2036 of consensus"
repeat_region
repeat_region
795. 1824
/note="L1M4 repeat: matches 4573. .5641 of consensus"
repeat_region
1824. .2482
/note="L1M4c repeat: matches 551. .1236 of consensus"
repeat_region
2469. 2811
/note="L1M4 repeat: matches 2332. .2692 of consensus"

```


AX321258
LOCUS AX321258 395 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 275 from Patent WO0177168.
ACCESSION AX321258
VERSION AX321258.1 GI:17904913
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
TITLE Compositions and methods for the therapy and diagnosis of lung cancer
JOURNAL
PATENT: WO 0177168-A 275 18-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1. .395
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 86 a 138 c 126 g 44 t 1 others
ORIGIN
Query Match 5.3%; Score 89.4; DB 6; Length 395;
Best Local Similarity 98.9%; Pred. No. 3.1e-15;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1608 AGCGGAGCGGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGGCGCT 1667
|||||
Db 13 AGGGGAGCGGAGAGCGGAGCCCGGAGAGAGCCCTGAGCAGCCCGCCGCGGCGCT 72
QY 1668 AGTTACCATCACACCCCGGGAGGAGCGGCAG 1698
|||||
Db 73 AGTTACCATCACACCCCGGGAGGAGCGGCAG 103
RESULT 12
AR117695
LOCUS AR117695 1481 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 3 from patent US 6140126.
ACCESSION AR117695
VERSION AR117695.1 GI:14098601
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1481)
AUTHORS Bennett,C.Frank. and Cowser,T.L.M.
TITLE Antisense modulation of Y-box binding protein 1 expression
JOURNAL Patent: US 6140126-A 3 31-OCT-2000;
FEATURES Location/Qualifiers
source 1. .1481
/organism="unknown"
BASE COUNT 437 a 381 c 386 g 277 t
ORIGIN
Query Match 5.2%; Score 89; DB 6; Length 1481;
Best Local Similarity 100.0%; Pred. No. 5.8e-15;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1610 CGGAGCGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGGCGCTAG 1669
Db 2 CGGAGCGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGGCGCTAG 61
QY 1670 TTACCATCACACCCCGGGAGGAGCGGCAG 1698
|||||
Db 62 TTACCATCACACCCCGGGAGGAGCGGCAG 90
RESULT 13
AX409453
LOCUS AX409453 1481 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2100 from Patent WO0229103.

AX409453
LOCUS AX409453.1 GI:21442158
DEFINITION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2100 11-APR-2002;
GENE LOGIC INC (US)
FEATURES Location/Qualifiers
source 1. .1481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. J03827"
BASE COUNT 437 a 381 c 386 g 277 t
ORIGIN
Query Match 5.2%; Score 89; DB 6; Length 1481;
Best Local Similarity 100.0%; Pred. No. 5.8e-15;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1610 CGGAGCGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGGCGCTAG 1669
|||||
Db 2 CGGAGCGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGGCGCTAG 61
QY 1670 TTACCATCACACCCCGGGAGGAGCGGCAG 1698
|||||
Db 62 TTACCATCACACCCCGGGAGGAGCGGCAG 90
RESULT 14
HUMYB1A
LOCUS HUMYB1A 1481 bp mRNA linear PRI 14-JAN-1995
DEFINITION Y box binding protein-1 (YB-1) mRNA.
ACCESSION J03827
VERSION J03827.1 GI:340418
KEYWORDS Y box binding protein.
SOURCE Human lymphoblastoid cell line (Swei) cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1481)
AUTHORS Didier,D.K., Schiftenbauer,J., Woulfe,S.L., Zacheis,M. and Schwartz,B.D.
TITLE Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (19), 7322-7326 (1988)
MEDLINE 89017190
PUBMED 3174636
COMMENT Draft entry and clean copy of sequence [1] kindly provided by B.D.Schwartz 31-AUG-1988.
FEATURES Location/Qualifiers
source 1. .1481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Unassigned"
1. .1481
/gene="NFYA"
127. 1080
/gene="NFYA"
/note="Y box binding protein-1"
/codon_start=1
/protein_id="AAA61308.1"
/db_xref="GI:340419"
/db_xref="GDB:G00-128-786"
/translation="MSSRETQPPAAAPPAALSAADTKPCTGSGAGSGPGGLTS
AAPAGDKRVIAIKVLGVTKVFNWNGYGF INRNDTKDFVHQTAIKKNPKYLR
VGDETVDFVDVGEKGBEAANYTGPVPGVSGKYAADRNHYRYPRRGPRNYQQ
NTQNSSEKNEGSAPGQAQRRPYRMRPYGRPPQYSNPPVQGEVM
ECADNQGAGEQGRPVQNNYRCYRPRFRGRPPRQRPREDGNEEDKENQDGTQSQQP

BASE COUNT 437 a 381 c 386 g 277 t
ORIGIN 2 bp upstream of NC11 cut site.
PORRYRRNRYRRRPNPKPODGKNETKRAADPPAENSRSG"

Query Match 5.2%; Score 89; DB 9; Length 1481;
Best Local Similarity 100.0%; Pred. No. 5.8e-15;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1610 CGGAGCGGAGACGCCACCCAGAGCCCTGAGCAGCCGCCACCGCGCGCGGCTAG 1669
|||||
Db 2 CGGAGCGGAGACGCCACCCAGAGCCCTGAGCAGCCGCCACCGCGCGCGGCTAG 61
|||||

QY 1670 TTACCATCACACCCCGGAGGAGCGGCAG 1698
|||||

Db 62 TTACCATCACACCCCGGAGGAGCGGCAG 90
|||||

RESULT 15
HUMPSDBPB 2168 bp DNA linear PRI 13-OCT-1995
LOCUS Homo sapiens DNA-binding protein B pseudogene (PSDBPB1) gene,
DEFINITION complete cds.

ACCESSION L37516
VERSION L37516.1 GI:950414
KEYWORDS DNA-binding protein B; pseudogene.
SOURCE Homo sapiens
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
AUTHORS Sakurai, H., Maekawa, T., Imamoto, F., Yasuda, K. and Ishii, S.
TITLE Two human genes isolated by a novel method encode DNA-binding
JOURNAL proteins containing a common region of homology
MEDLINE Eur. J. Biochem. 231 (1), 72-82 (1995)
PUBMED 95354705

REFERENCE 2 (bases 1 to 2168)
AUTHORS Kudo, S., Mattei, M.G. and Fukuda, M.
TITLE Characterization of the gene for dbpA, a family member of the
JOURNAL nucleic-acid-binding proteins containing a cold-shock domain
MEDLINE Eur. J. Biochem. 231 (1), 72-82 (1995)
PUBMED 7628487

FEATURES
source Location/Qualifiers

1..2168
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NAR12"
/tissue_type="placenta"
427..1978
/gene="PSDBPB1"
427..445
/evidence=experimental
/rpt_type=direct
581..1540
/gene="PSDBPB1"
/note="cold shock domain; RNP-1 motif"
/pseudo

/codon_start=1
/product="DNA-binding protein B"
1922..1927
/gene="PSDBPB1"
1960..1978
/evidence=experimental
/rpt_type=direct

BASE COUNT 664 a 531 c 510 g 463 t
ORIGIN

Query Match 5.1%; Score 87.4; DB 9; Length 2168;
Best Local Similarity 97.1%; Pred. No. 1.9e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1596 AGTTCGATCGGTACGGGAGCGGAGCGGACCCCGAGAGCCCTGAGCAGCCCGCCACCGC 1655
|||||

Db 444 AGTTCGATCGGTACGGGAGCGGAGCGGACCCCGAGAGCCCTGAGCAGCCCGCCACCGC 503
|||||

QY 1656 CGCCCGCGGCGCTAGTTACCATCATCACACCCCGGAGGAGCGGCAG 1698
|||||

Db 504 CG--GCCGCGCTAGTTACCGTCTACACCCCGGAGGAGCGGCAG 544
|||||

Search completed: November 24, 2002, 18:06:36
Job time : 5110 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2002, 15:01:58 ; Search time 348 Seconds
(without alignments)
10988.199 Million cell updates/sec

Title: HSYBL_COPY_453_2150

Perfect score: 1698

Sequence: 1 AGAACCATATGTCGCGCAA.....CACCCCGGAGGAGCCGCGAG 1698

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469.6	27.7	2102	23	AAS83003
2	111	6.5	178	21	Human secreted pro
3	107.4	6.3	1548	20	Human ovarian tumo
4	89.4	5.3	395	24	AAS61724
5	89	5.2	1481	22	Human small cell ca
6	88.6	5.2	1481	22	Human Y-box bindin
7	88.6	5.2	179	22	Gene #2100 used to
8	85.2	5.0	410	23	Human reproductive
9	83	4.9	3073	12	Human lung tumour- Clone associated w

10	83	4.9	3073	17	AAT34371	Plasmid pATG29 (AT
11	83	4.9	3073	20	AAZ32246	Human glioblastoma
12	83	4.9	3073	21	AA88181	pATG29 human gliob
13	81.4	4.8	1474	21	ABK84103	Human cDNA differe
14	78.8	4.6	381	21	AAC65939	Human lung cancer-
15	78.8	4.6	381	24	ABL49158	Human lung tumour
16	77.4	4.6	545	24	AAS61751	Lung small cell ca
17	61.2	3.6	2179	22	AAH15631	Human cDNA sequenc
18	47.4	2.8	60	24	ABN49420	Human spliced tran
19	45.2	2.7	3900	22	AAS26731	Human genomic DNA
20	45.2	2.7	3901	22	AAS26729	Human immunoglobul
21	44.8	2.6	3152	22	AAS28912	Sheep PrP gene for
22	41.8	2.5	4140	10	AAH92735	Mouse ischaemic co
23	41.6	2.4	1469	24	AB199233	Gene encoding a su
24	41.4	2.4	10732	21	AAAI0594	TNFR/NGER protein
25	41.2	2.4	2452	24	AAI71445	Human musculoskele
26	41	2.4	602	22	AAI37306	Clone of recombin
27	40.8	2.4	13104	14	AAO46852	Human chemically p
28	40.2	2.4	6731	24	ABK39863	Human immune syste
29	39.8	2.3	5195	24	ABL32920	Human DNA for stag
30	39.8	2.3	6944	21	ABK34026	Human ORFX ORF1824
31	38.8	2.3	584	21	AAC76269	Human DNA represen
32	38.8	2.3	98690	24	ABK12169	DNA encoding novel
33	38.4	2.3	160	22	AAS43317	DNA encoding novel
34	38.4	2.3	327	22	AAS43392	Human polynucleoti
35	38.4	2.3	327	22	AAI84918	Human cDNA sequenc
36	38.4	2.3	1952	22	AAH15912	Human cDNA sequenc
37	38.4	2.3	2287	22	AAH17021	Human breast speci
38	38.4	2.3	2399	24	ABT03034	Human single nucle
39	38	2.2	130	20	AAH86086	Human secreted pro
40	38	2.2	167	21	AAI21418	Colon adenocarcino
41	38	2.2	413	24	ABL62435	Ovary cancer relat
42	38	2.2	413	24	ABL67984	Human prostate exp
43	37.6	2.2	738	23	ABV23910	Human prostate exp
44	37.6	2.2	738	23	ABV29791	DNA transcription
45	37.2	2.2	6167	24	ABK28370	

ALIGNMENTS

RESULT 1

AAS83003
ID AAS83003 standard; cDNA; 2102 BP.

XX AC AAS83003;

DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #18807.

XX KW Human; chromosome mapping; gene mapping; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-0508631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG18816.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity -

PS Claim 1: SEQ ID No 18807; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2102 BP; 580 A; 540 C; 547 G; 434 T; 1 other;

Query Match 27.7%; Score 469.6; DB 23; Length 2102;
Best Local Similarity 97.8%; Pred. No. 2.3e-138;
Matches 497; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY 837 AGTAAAGAAAGATGCTGAATCTTGGAAAGCAAGACACATTTTTTAAATATATGCTGGT 896
DB 1 AGTAAAGAAAGATGCTGAATCTTGGAAAGCAAGACACATTTTTTAAATATATGCTGGT 60
QY 897 AAAACGGATCAGAGCAGGTCCCTCATGGAGCACACCTCGCCCTAAACATGCTGAACCC 956
DB 61 AAAACGGATCAGAGCAGGTCCCTCATGGAGCACACCTCGCCCTAAACATGCTGAACCC 120
QY 957 GGGCTGCATAGCTGGTGGTCCCTCCAAAGGTGACTGCTCCGACAAAAGGGTACGCTCT 1016
DB 121 GGGCTGCATAGCTGGTGGTCCCTCCAAAGGTGACTGCTCCGACAAAAGGGTACGCTCT 180
QY 1017 TCAACAGCA-TACGTTTAAGGCAATCCAGAAACCTCGGCTGCGCGCACTACACGGC 1075
DB 181 TCAACAGCACTAGTTAAGGCAATCCAGAAACCTCGGCTGCGCGCACTACACGGC 240
QY 1076 CATTAAGAAAGACGACTCTATGCGCGCGGTAAATGTTCTCAGATCACAGGACCGTATT 1135
DB 241 CATTAAGAAAGACGACTCTACGCGCGCGGTAAATGTTCTCAGATCACAGGACCGTATT 300
QY 1136 TGGAGCTGGAGGAGGAGGAGCCCTTTTCTTCACGGGGGGCTAAGGGCTCTTCGAGCCCCC 1195
DB 301 TGGAGCTGGAGGAGGAGGAGGAGCCCTTTTCTTCACGGGGGGTAAAGGGCTCTTCGAGCCCCC 360
QY 1196 TTCAATCCCGGTGCGCGGGGTAAATCCCTGCCAGCGTTCCGGGGTGCCTTTTCTCA 1255
DB 361 TTCAATCCCGGTGCGCGGGGTAAATCCCTGCCAGCGTTCCGGGGTGCCTTTTCTCA 420
QY 1256 GCCGAGACAAACCCCTGAAGCTGGGGGGCCCGCCAGCCCGCGGCTGCTCGTGGAAAGTCA 1315
DB 421 GCCGAGACTCAACCCCTGAAGCT-GGGGCCCGCCAGCCCGGAGCGCTGCTAGTGAAGTCA 479
QY 1316 CGTTCCTTCGCGCGCTCTCGGGTAC 1343
DB 480 CGTTCCTTCGCGCGCTCTCGGGTAC 507

RESULT 2
AAC09730

ID AAC09730 standard; cDNA; 178 BP.

XX AAC09730;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 13805.

XX Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1: SEQ ID 13805; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX Sequence 178 BP; 38 A; 70 C; 52 G; 17 T; 1 other;

Query Match 6.5%; Score 111; DB 21; Length 178;

Best Local Similarity 99.1%; Pred. No. 9.4e-25;

Matches 111; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1587 ATTCTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCCACAGAGCCCTGAGCAG 1646

DB 1 ATTCTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCCACAGAGCCCTGAGCAG 60

QY 1647 CCCACCGCGCGCGCGCTAGTTACCATCACACCCCGGAGGAGCCGCAG 1698

DB 61 CCCACCGCGCGCGCGCTAGTTACCATCACACCCCGGAGGAGCCGCAG 112

RESULT 3

AZ77507

ID AZ77507 standard; cDNA; 1548 BP.

XX AZ77507;

XX 10-APR-2000 (first entry)

XX Human ovarian tumor cDNA library derived EST fragment 58.

XX

KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW gene therapy; treatment; ss.
XX
OS Homo sapiens.

XX
PN DE19817557-Al.
XX
PD 21-OCT-1999.

XX
PF 09-APR-1998; 98DE-1017557.

XX
PR 09-APR-1998; 98DE-1017557.

XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarisky C, Dahl E;
XX WPI; 1999-591920/51.

XX
DR New nucleic acid sequences expressed in ovarian, and some other, cancer
XX tissues, and derived polypeptides, for treatment of ovarian cancer and
XX identification of therapeutic agents
XX
XX Claim 3; Page 186; 310pp; German.

XX
CC This invention describes novel nucleic acid (cDNA) sequences (A) which
CC have anticancer activity and are highly expressed in ovarian tumor
CC tissue (and some also in testis and breast cancer tissue). The products
CC of the invention can be used for gene therapy. (A) are used (i) for
CC recombinant expression of polypeptides (B) and (ii) to isolate complete
CC genes. (B) are used (i) to identify agents suitable for treatment of
CC ovarian cancer; (ii) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (iii) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AA277430-277572 represent the human
CC ovarian tumor cDNA library derived EST fragments described in the method
CC of the invention and encode the protein fragments represented in
CC AA76505-Y76638.

XX
SQ Sequence 1548 BP; 454 A; 386 C; 403 G; 305 T; 0 other;

Query Match 6.3%; Score 107.4; DB 20; Length 1548;
Best Local Similarity 99.1%; Pred. No. 5.4e-23;
Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1590 CTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCCCGGAGAGCCCTGAGCAGCC 1649

Db 1 CTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCCCGGAGAGCCCTGAGCAGCC 60

QY 1650 CACCGCCGCCCGCGGCTAGTTACCATCACACCCCGGAGGAGCGCGAG 1698

Db 61 CACCGCCGCCCGCGGCTAGTTACCATCACACCCCGGAGGAGCGCGAG 109

RESULT 4

AAS61724
ID AAS61724 standard; cDNA; 395 BP.

XX
AC AAS61724;

XX
DT 29-JAN-2002 (first entry)

XX
DE Lung small cell carcinoma antigen, cDNA #265.

XX
KW Human; cytostatic; antitumour; lung small cell cancer antigen;
XX tumour; lung cancer; ss.

XX
OS Homo sapiens.

XX
PN WO200177168-A2.

XX
PD 18-OCT-2001.

XX
PF 11-APR-2001; 2001WO-US11859.

XX
PR 11-APR-2000; 2000US-196780P.

XX
PR 21-JUN-2000; 2000US-213361P.

XX
PR 01-SEP-2000; 2000US-229763P.

XX
PR 05-SEP-2000; 2000US-230629P.

XX
PR 14-SEP-2000; 2000US-232565P.

XX
PR 19-DEC-2000; 2000US-257037P.

XX
PR 08-JAN-2001; 2001US-260796P.

XX
PA (CORI-) CORIXA CORP.

XX
PI Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX WPI; 2002-010896/01.

XX
DR Lung tumour polynucleotide and polypeptides useful in therapy and
XX diagnosis of cancer especially lung cancer
XX
XX Claim 1; Page 226; 295pp; English.

XX
CC The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by
CC utilising oligonucleotides (III), where the biological sample
CC from the patient is contacted with (III), detecting the amount of
CC polynucleotide hybridised to (III) in the sample and comparing the
CC amount of polynucleotide to a predetermined cut-off value and thereby
CC determining cancer in a patient. (I), (II) or antigen-presenting cells
CC expressing (II) is useful for stimulating and/or expanding T cells
CC specific for a tumour protein. The method comprises contacting T cells
CC with one of the components under conditions to permit the stimulation
CC and/or expansion of the cells. A composition comprising (I) is useful for
CC stimulating an immune response in a patient and for inhibiting the
CC development of a cancer especially lung cancer in a patient. An
CC isolated T cell population is useful for removing tumour cells from the
CC biological sample and for inhibiting the development of cancer in a
CC patient. AAS61460-AAS61874 represent novel human lung small cell
CC cancer antigen coding sequences of the invention.

XX
SQ Sequence 395 BP; 86 A; 138 C; 126 G; 44 T; 1 other;

Query Match 5.3%; Score 89.4; DB 24; Length 395;
Best Local Similarity 98.9%; Pred. No. 1.2e-17;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1608 AGCGGAGCGGAGCGGAGCCCGGAGAGCCCTGAGCAGCCCGGAGCGCGGCT 1667

Db 13 AGGGGAGCGGAGAGCGGAGCCCGGAGAGCCCTGAGCAGCCCGGAGCGCGGCT 72

QY 1668 AGTTACCATCACACCCCGGAGGAGCGCGAG 1698

Db 73 AGTTACCATCACACCCCGGAGGAGCGCGAG 103

RESULT 5

AAC81319
ID AAC81319 standard; DNA; 1481 BP.

XX
AC AAC81319;

XX
DT 23-FEB-2001 (first entry)

XX
DE Human Y-box binding protein 1 DNA.

XX
KW Human Y-box binding protein 1; YB-1; DNA binding protein B; dbpB;
KW transcription factor; nucleic acid binding; DNA repair;

XX
KW cell sensitisation; genotoxic stress; immune regulation; MHC expression;

KW viral gene expression; extracellular matrix degradation regulator;
KW redox signalling; expression inhibition; antisense therapy;
KW tumour formation; cancer multidrug resistance; inflammation;
KW immune disorder; infection; ds.
XX
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 127..1080
FT FT /*tag= a
FT FT /product= "Human Y-box binding protein"
FT FT /function= "Transcription factor"
FT FT misc_feature 1032..1077
FT FT /*tag= b
FT FT /note= "The 15 amino acids encoded by this
FT FT region are not included in the
FT FT corresponding protein sequence"
XX
XX
XX US6140126-A.
XX 31-OCT-2000.
XX 26-OCT-1999; 99US-0429323.
XX 26-OCT-1999; 99US-0429323.
XX (ISIS-) ISIS PHARM INC.
XX Bennett CF, Cowseart LM;
XX WPI: 2001-0233284/03.
XX P-PSDB; AAB29744.
XX Antisense oligonucleotides, useful for modulating the expression of
FT Y-box binding protein 1, as well as for treating or preventing diseases
FT associated with Y-box binding protein 1 expression, e.g. inflammation
FT or tumor formation
XX
XX Example 15; Column 47-50 ; 40pp; English.
XX This sequence represents DNA encoding human Y-box binding protein 1.
CC Human Y-box binding protein 1 (also known as YB-1, DNA binding protein B
CC and dbpB) is a member of the Y-box binding protein family of
CC transcription factors, a highly conserved family of nucleic acid binding
CC proteins which bind to the Y-box, an inverted CCAAT sequence found in
CC the promoters of many genes. Y box binding proteins have a broad
CC specificity for nucleic acids, being able to bind double-stranded DNA,
CC damaged DNA, and single-stranded DNA and RNA. Y-box binding protein 1
CC plays a role in DNA repair and the sensitisation of cells from a diverse
CC array of genotoxic stresses, including DNA cross-linking agents and
CC ultraviolet irradiation. Y-box binding protein 1 is also involved in
CC immune regulation, being a negative regulator of MHC (major
CC histocompatibility complex) gene expression, and additionally modulates
CC viral gene expression. It also participates in the regulation of
CC extracellular matrix degradation, and is thought to be involved in redox
CC signalling. The invention relates to antisense oligonucleotides targeted
CC to the human Y-box binding protein 1 gene, which inhibit its expression.
CC A series of oligonucleotides (AAC81326-C81405) were designed to target
CC different regions of the human Y-box binding protein 1 mRNA, and were
CC analysed for their effect on Y-box binding protein 1 mRNA levels by
CC quantitative real-time PCR. The oligonucleotides of the invention are
CC useful for diagnosis, prevention and treatment of conditions associated
CC with Y-box binding protein 1 expression, such as tumour formation,
CC cancer multidrug resistance, inflammation, immune disorders and certain
CC infections.
XX
XX Sequence 1481 BP: 437 A; 381 C; 386 G; 277 T; 0 other;
Query Match 5.2%; Score 89; DB 22; Length 1481;
Best Local Similarity 100.0%; Pred. No. 3.8e-17;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1610 CGGAGCGGAGCGGACCCAGAGAGCCCTGAGAGCCCCACCGCGCGGCGCTAG 1669

Db 2 CGGAGCGGAGCGGACCCAGAGAGCCCTGAGAGCCCCACCGCGCGGCGCTAG 61
QY 1670 TTACCATCATCACCCCGGAGGAGCGCGAG 1698
Db 62 TTACCATCATCACCCCGGAGGAGCGCGAG 90
RESULT 6
ABN95602
ID ABN95602 standard; DNA; 1481 BP.
XX
XX AC ABN95602;
XX
XX DT 13-AUG-2002 (first entry)
XX
XX DE Gene #2100 used to diagnose liver cancer.
XX
XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
XX OS Homo sapiens.
XX
XX PN WO200229103-A2.
XX
XX PD 11-APR-2002.
XX
XX PF 02-OCT-2001; 2001WO-US30589.
XX
XX PR 02-OCT-2000; 2000US-237054P.
XX
XX PA (GENE-) GENE LOGIC INC.
XX
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI: 2002-426119/45.
XX
XX PT Diagnosing and detecting the progression of liver cancer.
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient.
XX PT involves detecting the level of expression of two or more genes in a
XX liver tissue sample
XX
XX PS Claim 1; SEQ ID NO 2100; 298pp; English.
XX
XX CC The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumour in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1481 BP: 437 A; 381 C; 386 G; 277 T; 0 other;
Query Match 5.2%; Score 89; DB 24; Length 1481;
Best Local Similarity 100.0%; Pred. No. 3.8e-17;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1610 CGGAGCGGAGCGGACCCAGAGAGCCCTGAGAGCCCCACCGCGCGGCGCTAG 1669
Db 2 CGGAGCGGAGCGGACCCAGAGAGCCCTGAGAGCCCCACCGCGCGGCGCTAG 61
QY 1670 TTACCATCATCACCCCGGAGGAGCGCGAG 1698

PT and cAMP, that modify, complement or suppress genetic defect and
 PT correct associated phenotypic alteration
 XX
 PS Disclosure; Page 111; 169pp; English.
 XX
 CC In the specification this sequence is given the SEQ ID NO. 30.
 CC There is no reference to SEQ ID NO.30 in the text. Other sequences
 CC in the specification were isolated from human glioblastoma cells
 CC and encode cyclic nucleotide PDEs and RAS-related polypeptides.
 CC (They were isolated by their ability to complement or suppress
 CC genetic defects in a biochemical pathway involving cAMP or which is
 CC controlled by a RAS protein).
 XX
 SQ Sequence 3073 BP; 897 A; 684 C; 635 G; 857 T; 0 other;
 Query Match 4.9%; Score 83; DB 12; Length 3073;
 Best Local Similarity 100.0%; Pred. No. 5e-15;
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 1675
 Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 78
 QY 1676 TCACACCCCGGAGAGCCGCGAG 1698
 Db 79 TCACACCCCGGAGAGCCGCGAG 101
 RESULT 10
 AAT34371
 ID AAT34371 standard; cDNA; 3073 BP.
 XX
 AC AAT34371;
 XX
 DT 09-OCT-1996 (first entry)
 XX
 DE Plasmid pATG29 (ATCC 68591) insert.
 XX
 KW Human; glioblastoma; complementation; S. cerevisiae; S. pombe;
 KW Clone S46; strain RS60.15B; RAS2; RAS2(vall19ala15); Xenopus laevis;
 KW S6 protein kinase; Plasmid pML5; strain SRN37; CAP; adenylyl cyclase;
 KW diploid; strain SP565; ras1; ras1::LEU2; mutation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..1112
 FT /*tag= a
 FT /product= RAS-related protein
 XX
 XX US5527896-A.
 XX
 XX 18-JUN-1996.
 XX
 XX 20-APR-1990; 90US-0511715.
 XX
 XX 19-APR-1991; 91US-0688352.
 XX 20-APR-1990; 90US-0511715.
 XX
 XX (COLD-) COLD SPRING HARBOR LAB.
 XX
 XX Colicelli JJ, Wigler MH;
 XX
 XX WPI; 1996-299902/30.
 XX P-PSDB; AAW00089.
 XX
 XX DNA mols. isolated from human glioblastoma cells - encode
 XX RAS-related or cyclic nucleotide phosphodiesterase proteins
 XX
 XX Claim 1; Column 107-112; 101pp; English.
 XX
 XX The sequences given in AAT34369-71 and AAT34391 represent plasmid
 XX fragments which contain human glioblastoma cell cDNA inserts which do not
 CC

CC encode phosphodiesterases. These cDNA's were obtained by complementation
 CC of two genetically altered S. cerevisiae and S. pombe strains. Clone
 CC S46 was selected by complementation in S. cerevisiae strain RS60.15B.
 CC This strain contains a mutant allele of RAS2, RAS2(vall19ala15), which
 CC renders cells unable to grow at 36 deg. C, because such cells are
 CC defective in RAS function at elevated temperatures. Human cDNA's from a
 CC human glioblastoma cell library were selected that could complement this
 CC defect. The deduced sequence of S46 is homologous to a Xenopus laevis
 CC gene that encodes a known protein kinase, the S6 protein kinase.
 CC Plasmid pML5 was selected by complementation in S. cerevisiae strain
 CC SRN37. This strain contains a disrupted allele of CAP, cap::HIS3. CAP
 CC encodes an adenylyl cyclase associated protein of undetermined function.
 CC As a consequence of this gene disruption, SRN37 fails to grow in medium
 CC rich in amino acids. Human cDNA's were chosen which could complement
 CC this defect. Plasmids pATG16 and pATG29 were selected by
 CC complementation in the S. pombe diploid strain SP565. This strain is
 CC homozygous for disruptions of ras1 (ras1::LEU2). As a result of this
 CC mutation, this strain fails to sporulate, and human cDNA's were selected
 CC which could complement this defect. These genes have unknown function.
 XX
 SQ Sequence 3073 BP; 897 A; 684 C; 635 G; 857 T; 0 other;
 Query Match 4.9%; Score 83; DB 17; Length 3073;
 Best Local Similarity 100.0%; Pred. No. 5e-15;
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 1675
 Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 78
 QY 1676 TCACACCCCGGAGAGCCGCGAG 1698
 Db 79 TCACACCCCGGAGAGCCGCGAG 101
 RESULT 11
 AAZ32246
 ID AAZ32246 standard; cDNA; 3073 BP.
 XX
 AC AAZ32246;
 XX
 DT 19-JAN-2000 (first entry)
 XX
 DE Human glioblastoma cell RAS-related pATG29 encoding cDNA.
 XX
 KW Phosphodiesterase; dunce-like phosphodiesterase; PDE; DPD; cAMP;
 KW RAS-related protein; immunoreactive; detection; genetic defect;
 KW bronchodilation; increased myocardial contractility;
 KW anti-inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 XX US5977305-A.
 XX
 XX 02-NOV-1999.
 XX
 XX 07-JUN-1995; 95US-0474379.
 XX
 XX 01-MAR-1994; 94US-0206188.
 XX 20-APR-1990; 90US-0511715.
 XX 19-APR-1991; 91US-0688352.
 XX
 XX (COLD-) COLD SPRING HARBOR LAB.
 XX
 XX Colicelli JJ, Wigler MH;
 XX
 XX WPI; 1999-619709/53.
 XX P-PSDB; AAY49814.
 XX
 XX New isolated RAS-related polypeptides and mammalian cyclic nucleotide
 XX phosphodiesterases, used for screening for agents which can modify
 XX complement or suppress genetic defects -
 XX

PS Example 1; Column 115-120; 145pp; English.

XX The present invention describes new isolated RAS-related polypeptides

CC and mammalian cyclic nucleotide phosphodiesterases (PDEs). RAS-related

CC polypeptides are capable of complementing a defective RAS function in

CC yeast. The products can be used for screening for agents which can

CC modify, complement or suppress a genetic defect in a biochemical

CC pathway in which cAMP participates, or in a biochemical pathway which

CC is controlled, directly or indirectly, by a RAS protein and other

CC proteins affecting cell growth and maintenance. Developing agents that

CC will selectively act upon PDEs is directed toward reproducing the

CC desirable effects of cyclic nucleotides, e.g. bronchodilation,

CC increased myocardial contractility, anti-inflammation, yet without

CC causing the undesirable effects, e.g. increased heart rate or enhanced

CC lipolysis. The products can also be used for therapeutic, diagnostic

CC and prognostic uses. AA232229 to AA232285, and AA49803 to AA49830,

CC represent sequences used in the exemplification of the present

CC invention.

XX SQ Sequence 3073 BP; 897 A; 684 C; 635 G; 857 T; 0 other;

Query Match 4.9%; Score 83; DB 20; Length 3073;

Best Local Similarity 100.0%; Pred. No. 5e-15;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 1675

Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 78

QY 1676 TCACACCCCGGAGGAGCGCGCAG 1698

Db 79 TCACACCCCGGAGGAGCGCGCAG 101

RESULT 12

AAA88181

ID AAA88181 standard; cDNA; 3073 BP.

XX AC AAA88181;

XX 14-DEC-2000 (first entry)

DT PATG29 human glioblastoma cell insert nucleotide sequence SEQ ID NO:31.

DE Detection; mammalian gene; yeast; microorganism; identification;

XX phenotype; characteristic; dunce-like phosphodiesterase; PDE; RAS;

KW RAS-related protein; genetic defect; hybridisation; probe; ss.

XX Homo sapiens.

OS Schizosaccharomyces pombe.

XX US6100025-A.

XX 08-AUG-2000.

XX 01-MAR-1994; 94US-0206188.

XX 20-APR-1990; 90US-0511715.

PR 19-APR-1991; 91US-0688352.

XX (COLD-) COLD SPRING HARBOR LAB.

PA Colicelli JJ, Wigler MH;

PI WPI; 2000-531664/48.

DR P-PSDB; AAB20625.

XX Novel isolated DNA encoding a mammalian cyclic nucleotide

PT phosphodiesterase is present in plasmids pPDE46, pPDE43 or pPDE339 and

PT is used to modify a genetic defect in a biochemical pathway in which

PT cAMP participates

XX Example 1; Column 121-126; 145pp; English.

PS

XX The present invention describes a purified and isolated DNA (I) which

CC encodes a mammalian cyclic nucleotide phosphodiesterase and is an insert

CC present in the plasmids pPDE46 (ATCC 69552), pPDE43 (ATCC 69551) or

CC pPDE339 (ATCC 69550). The DNA molecules are used to modify, complement

CC or suppress a genetic defect in a biochemical pathway in which cAMP

CC participates and are also used as hybridisation probes. The present

CC invention also describes methods for detecting mammalian genes encoding

CC proteins which can function in microorganisms, particularly yeast, to

CC modify, complement, or suppress a genetic defect associated with an

CC identifiable phenotypic alteration or characteristic in the

CC microorganism. AAA88162 to AAA88218 and AAB29614 to AAB20640 represent

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 3073 BP; 897 A; 684 C; 635 G; 857 T; 0 other;

Query Match 4.9%; Score 83; DB 21; Length 3073;

Best Local Similarity 100.0%; Pred. No. 5e-15;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 1675

Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGCTAGTTACCA 78

QY 1676 TCACACCCCGGAGGAGCGCGCAG 1698

Db 79 TCACACCCCGGAGGAGCGCGCAG 101

RESULT 13

ABK84103

ID ABK84103 standard; cDNA; 1474 BP.

XX AC ABK84103;

XX 14-AUG-2002 (first entry)

DT Human cDNA differentially expressed in granulocytic cells #674.

DE Human; ss; granulocytic cell; DNA chip; bacterial infection;

XX viral infection; parasitic infection; protozoal infection;

KW fungal infection; sterile inflammatory disease; psoriasis;

KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;

KW cardiac reperfusion injury; renal reperfusion injury; ARDS;

KW adult respiratory distress syndrome; inflammatory bowel disease;

KW Crohn's disease; ulcerative colitis; periodontal disease;

KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

OS WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

XX 03-OCT-2000; 2000US-237189P.

PA (GENE-) GENE LOGIC INC.

XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

PI WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression

PT of genes associated with granulocyte activation, which serves as

PT diagnostic markers that is useful for monitoring disease states and

PT drug toxicity

XX Claim 1; SEQ ID No 674; 114pp; English.

PS The invention relates to detecting (M1) granulocyte (GC) activation

CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by

CC

XX Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;
XX Vedvick TS, Carter D, Watanabe Y, Peckham DW;

DR WPI; 2002-090513/12.

XX Polynucleotides encoding lung tumor polypeptides, useful for treating
PT lung cancer or stimulating an immune response -

XX Example 1; Page 279; 374pp; English.

XX The present invention describes human lung tumour proteins. Human lung
CC tumour proteins and polynucleotides have cytostatic and immunostimulant
CC activities, and can be used in vaccine production. Compositions
CC comprising the lung tumour proteins, polynucleotides, antibodies,
CC fusion proteins, T cell populations, or antigen presenting cells that
CC express the lung tumour proteins are useful for treating lung cancer or
CC stimulating an immune response. ABL48959 to ABL49100 and ABB74946 to
CC ABB75070 represent sequences used in the exemplification of the present
CC invention.

XX SQ Sequence 381 BP; 83 A; 135 C; 117 G; 41 T; 5 other;

Query Match 4.6%; Score 78.8; DB 24; Length 381;
Best Local Similarity 93.0%; Pred. No. 2.8e-14;
Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1613 GAGCGGAGAGCGGACCCAGAGACCCCTGTGACACGCCCGCGCGGCTAGTTA 1672

Db 1 GAGCGGAGAGCGGACCCAGAGACCCCTGTGACACGCCCGCGCGGCTAGTTN 60

Qy 1673 CCATCACACCCCGGAGAGCGCGCAG 1698

Db 61 NCATCACACCCCGGAGAGCGCGCAG 86

Search completed: November 24, 2002, 16:41:50
Job time : 385 secs

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-07-688-352C-31

Query Match 4.9%; Score 83; DB 1; Length 3073;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGGAGAGCGGACCCCGAGAGAGCCCTGAGCAGCCCCCAGCCCGCGCGCTAGTTACCA 1675
|||||
Db 19 CGGAGAGCGGACCCCGAGAGAGCCCTGAGCAGCCCCCAGCCCGCGCGCTAGTTACCA 78
|||||

QY 1676 TCACACCCCGGAGGAGCCGCGAG 1698
|||||
Db 79 TCACACCCCGGAGGAGCCGCGAG 101
|||||

RESULT 3
US-08-474-379C-31
Sequence 31, Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
TITLE OF INVENTION: PROCESSES
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,379C

FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,188
FILING DATE: 01-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,352
FILING DATE: 19-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 27866/32771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1109
US-08-474-379C-31

Query Match 4.9%; Score 83; DB 2; Length 3073;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGGAGAGCGGACCCCGAGAGAGCCCTGAGCAGCCCCCAGCCCGCGCGCTAGTTACCA 1675
|||||
Db 19 CGGAGAGCGGACCCCGAGAGAGCCCTGAGCAGCCCCCAGCCCGCGCGCTAGTTACCA 78
|||||

QY 1676 TCACACCCCGGAGGAGCCGCGAG 1698
|||||
Db 79 TCACACCCCGGAGGAGCCGCGAG 101
|||||

RESULT 4
US-09-146-249A-31
Sequence 31, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-09-146-249A-31

Query Match 4.9% Score 83; DB 3; Length 3073;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGGCTAGTTACCA 1675
Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGCGGCTAGTTACCA 78
Qy 1676 TCACACCCCGGAGGAGCGCGAG 1698
Db 79 TCACACCCCGGAGGAGCGCGAG 101

RESULT 5
US-08-206-188B-31
Sequence 31, Application US/08206188B
Patent No. 6100025
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60605-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,188B
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-08-206-188B-31

Query Match 4.9% Score 83; DB 3; Length 3073;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGGCTAGTTACCA 1675
Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCGCGGCTAGTTACCA 78
Qy 1676 TCACACCCCGGAGGAGCGCGAG 1698
Db 79 TCACACCCCGGAGGAGCGCGAG 101

RESULT 6
PCT-US91-02714-30
Sequence 30, Application PC/TUS9102714
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02714
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
PCT-US91-02714-30

Query Match 4.9% Score 83; DB 5; Length 3073;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: DORNEE

*
*

*
*
*
*
*

*
*
*
*
*

*
*
*
*
*

*
*

GENERAL INFORMATION:
APPLICANT: DORNER, F.

APPLICANT

Db 1408 YYYYYYYYYYYYYYYYYYYYYY 1434

RESULT 9

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOMLOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 2.6%; Score 44.2; DB 1; Length 7218;

Best Local Similarity 3.8%; Pred. No. 0.002;

Matches 13; Conservative 192; Mismatches 140; Indels 0; Gaps 0;

QY 597 AGAAGGAGAAAGTCTGCTGGGAGGACTTCAGTAACATCAGGTGGCGAGCTTAAAAA 656

Db 1391 RRR 1332

QY 657 AGAAGGAGAAAGTCTGCTGGGAGGACTTCAGTAACATCAGGTGGCGAGCTTAAAAA 716

Db 1331 RRR 1272

QY 717 CGTTTGAAACGCTAGTAGTAATCCCTCTATCAGGTGGCTGTTCAGAGATAAGTAA 776

Db 1271 RRR 1212

QY 777 AAAACAAACAGCTAGCTTCTCAATAAATGTCAGTTGAATTAATCTGATTGTGCTC 836

Db 1211 RRR 1152

QY 837 AGTAGAAAAAGATGTGAATCTTGGAAAGAACACATTTTTTTAAATATATATCCCTGGT 896

Db 1151 RRR 1092

QY 897 AAAACGATCAGAGGAGGAGTCCCATCGAGACACACCTCGCCCT 941

Db 1091 RRR 1047

RESULT 10

US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 2.6%; Score 43.4; DB 4; Length 289;

Best Local Similarity 8.2%; Pred. No. 0.00043;

Matches 20; Conservative 104; Mismatches 119; Indels 0; Gaps 0;

QY 223 TGCATTAGCCTCCTAAGTCTTCTCTACTTCCCTCTGCGCCCTCTGCAACCCATTCT 282

Db 243 YGYTTYAYCYCYAYCYCYTYGYSYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYN 184

QY 283 CCATCCGCGAGCCATTTTAAAAAGATGCCCTCCCTCTACTTATGACTCTAAATGTCT 342

Db 183 YSYNY 124

QY 343 TCTCATTCTCCCTCAGGATATATTCCCAATTAATATACCTAAGTACTGCCACCTC 402

Db 123 YSYNY 64

QY 403 TGAACCCCAATGTACATTCGAGTCTTACTGACTTCTGACTTCATTTCCCGAGATCTC 462

Db 63 YSYNYNYCYATYTYGYTYAYAYTYTYGYTYAYAYTYTYGYTYAYAYTYTYGYTCYC 4

QY 463 ACC 465

Db 3 YCY 1

RESULT 11

US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN


```

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 19250 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-35

```

	Query Match	2.0%	Score 34.6;	DB 4;	Length 19250;
	Best Local Similarity	50.3%;	Pred. No. 5;		
	Matches	85;	Conservative	0;	Mismatches 84; Indels 0; Gaps 0;
QY	733	ATAGTAATCCCTCTATACAGCTGGCTGTTGCAGGAAATAAGTGAATAAAACAAAACAGGCTA	792		
Db	13408	ATATPTCTAAATTAGTATGTTATCTATTAGAGACAAAGAGAGCTTAGGAAGAAGAGCTC	13349		
QY	793	GCTTGTTCATTAATGTGAGTTGAATTAATCTCATTTTGTGCTCAGTAGAAAAAGATGTG	852		
Db	13348	GTGTGTTACTATGAATACTCTTTCAANAGAGCACTTTATGCTCGAACTGGAGACATGTT	13289		
QY	853	AACTACTTGGAAAGGAGACACATTTTTTTAAATATATGCCGTGTAAC	901		
Db	13288	TAGAGAGGGGAAGTAAAGAAACAATAATTTATGAGAGTACTGTTTATAC	13240		

Search completed: November 24, 2002, 18:36:06
Job time : 276 secs

; APPLICANT: Kalos, Michael D.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Retter, Marc W
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY.
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (17)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (20)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (60)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (61)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (365)
; OTHER INFORMATION: n=A,T,C or G
; US-09-850-716A-215

Query Match 4.6%; Score 78.8; DB 10; Length 381;
Best Local Similarity 93.0%; Pred. No. 5.5e-15;
Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1613 GAGCGGAGCGGAGCCCTGAGCAGCCCTGAGCAGCCCGCGCGCGGCTAGTTA 1672
|||||
Db 1 GAGCGGAGCGGAGCCCTGAGCAGCCCTGAGCAGCCCGCGCGCGGCTAGTTN 60

QY 1673 CCATCACACCCCGGAGGAGCGCGAG 1698
|||||
Db 61 NCATCACACCCCGGAGGAGCGCGAG 86

RESULT 6
US-09-897-778-215
; Sequence 215, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 17, 20, 60, 61, 365

; OTHER INFORMATION: n = A,T,C or G
US-09-897-778-215

Query Match 4.6%; Score 78.8; DB 10; Length 381;
Best Local Similarity 93.0%; Pred. No. 5.5e-15;
Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1613 GAGCGGAGCGGAGCCCGGAGGAGCCCTGAGCAGCCCGCGCGCGGCTAGTTA 1672
|||||
Db 1 GAGCGGAGCGGAGCCCTGAGCAGCCCGCGCGCGGCTAGTTN 60

QY 1673 CCATCACACCCCGGAGGAGCGCGAG 1698
|||||
Db 61 NCATCACACCCCGGAGGAGCGCGAG 86

RESULT 7
US-09-833-790-302
; Sequence 302, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 302
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-833-790-302

Query Match 4.6%; Score 77.4; DB 10; Length 545;
Best Local Similarity 98.7%; Pred. No. 2e-14;
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1620 GAGCGGAGCCCGGAGGAGCCCTGAGCAGCCCGCGCGCGGCTAGTTACATCAC 1679
|||||
Db 13 GAGCGGAGCCCGGAGGAGCCCTGAGCAGCCCGCGCGCGGCTAGTTACATCAC 72

QY 1680 ACCCGGAGGAGCGCGAG 1698
|||||
Db 73 ACCCGGAGGAGCGCGAG 91

RESULT 8
US-09-764-864-1705/c
; Sequence 1705, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1705
; LENGTH: 3900
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-864-1705

Query Match 2.7%; Score 45.2; DB 10; Length 3900;

Best Local Similarity 72.0%; Pred. No. 0.0015;
Matches 59; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

[illegible]

RESULT 9

```

US-09-764-864-1703
; Sequence 1703, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1703
; LENGTH: 3901
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1703

```

Query Match 2.7%; Score 45.2; DB 10; Length 3901;
Best Local Similarity 72.0%; Pred. NO. 0.0015;
Matches 59; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

[illegible]

RESULT 10

```

RESOLVING ID
US-09-764-877-3671/c
; Sequence 3671, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3671
; LENGTH: 602.
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3671

```

Query Match 2.4%; Score 41; DB 10; Length 602;
Best Local Similarity 69.1%; Pred. No. 0.0088;
Matches 56; Conservative 0; Mismatches 25; Indels 0; Gaps 0.

[illegible]

Db 125 CTAATCAAGTAAATAAAAAA 105

RESULT 11

```

US-10-001-887-39/c
: Sequence 39, Application US/10001887
: Patent No. US20020155464A1
: GENERAL INFORMATION:
: APPLICANT: Salceda, Susana
: APPLICANT: Macina, Roberto
: APPLICANT: Recipon, Hervé
: APPLICANT: Cafferey, Robert
: APPLICANT: Sun, Yongming
: APPLICANT: Liu, Chenghua
: TITLE OF INVENTION: Compositions and
: FILE REFERENCE: DEX-0269
: CURRENT APPLICATION NUMBER: US/10/0
: CURRENT FILING DATE: 2001-11-20
: PRIOR APPLICATION NUMBER: 60/249,99
: PRIOR FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: 60/252,56
: PRIOR FILING DATE: 2000-11-22
: NUMBER OF SEQ ID NOS: 137
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 39
: LENGTH: 2399
: TYPE: DNA
: ORGANISM: Homo sapien
US-10-001-887-39

```

Query Match 2.3%; Score 38.4; DB 9; Length 2399;
Best Local Similarity 67.5%; Pred. NO. 0.16;
Matches 54; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

[illegible]

RESULT 12

```

RES001 12
US-09-960-352-2899
; Sequence 2899, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 2899
; LENGTH: 280
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 13-LIB3058-009-Q1-K1-D9
US-09-960-352-2899

```

Query Match 2.2%; Score 38; DB 10; Length 280;
Best Local Similarity 51.8%; Pred. No. 0.046;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy	652	AAAAAAGAACGGAGAAGGTGGTGGGAGGACTTCAGTAACATCAGGTGGCGACCTCAAT	711
Db	97	AGAAATAGCTTTAAGAGGAGTGAAAGATGACTCTTGTGAAGATATTGCGAAATCTTCAAT	156


```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007462.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
US-09-864-761-9089

Query Match      2.1%; Score 36.2; DB 10; Length 562;
Best Local Similarity 46.9%; Pred.No. 0.29;
Matches 145; Conservative 0; Mismatches 163; Indels 1; Gaps 1;

QY 352 TCCCTCAGGATATATTCACAATAAATATACCTAAGTGACTGCCACCTCTGCAACCCA 411
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82 TACCATCTGCTTTTATTTCTGGAAGTTGTCTCAGTATCAGGCTACTGACCTCTGCAA-CCT 140

QY 412 ATGTCACATTCGAGTCTTACTGAACACTACTTGACTGCTATTCGCGAGATCTACCTCTTCT 471
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 141 ATTTATCATTTGCATACAAAATATTCCTTTCTTCTTCCCTCTCCCTTTCTCCTCC 200

QY 472 CGCTGTGTACCTCTGCGGGAAGTCAGCCCTCCACCTTCTCCCTGCTTCCACTCCCAAA 531
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 201 CTCTTCCCCTCCCTTCCACCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTTCTCTCA 260

QY 532 ATACTTCGTGGTTTTCAGCTCTGGAGTATTTACCGTGTGGCTGTTAAATTTCTGCGCT 591
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 CTCCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 320

QY 592 CCATCAGAAGGAGAACTGACTCGGCAACTATTCATCCCGCGCGATAGTAGACGCTT 651
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 GATACATATCTTCTACTACGTTTTCGCTGCCCAATACTCCCTGGGAAGAAATCAGAAAAA 380

QY 652 AAAAAAGAA 660
   ||||| |||||
Db 381 AAAAAAAA 389
```

Search completed: November 24, 2002, 18:35:49
Job time : 165 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2002, 16:33:38 ; Search time 2218 seconds
(without alignments)
12398.529 Million cell updates/sec

Title: HSYBL_COPY_453_2150

Perfect score: 1698

Sequence: 1 AGAACATATGTCGGCCAA.....CACCCGGGAGGAGCGCCAG 1698

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estlin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_hic:**

9: gb_estl:**

10: gb_est2:**

11: gb_hic:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	519.4	30.6	532	9	AI659243
C 2	489.4	28.8	503	13	BI830189
C 3	451.6	26.6	565	12	BF111618
C 4	436	25.7	509	9	AA463628
C 5	425.4	25.1	438	9	AI650982
C 6	418.4	24.6	431	9	AI268811
					AI659243 tul8b06.x
					BI830189 603072855
					BF111618 7132c03.x
					AA463628 zx99g06.s
					AI650982 wa96d06.x
					AI268811 qo40g10.x

C 7	410.4	24.2	428	9	AI467820
C 8	405.8	23.9	428	9	AI208001
C 9	399.4	23.5	412	14	N67996
C 10	354.4	20.9	371	10	AM448961
C 11	354.2	20.8	378	12	BE58954
C 12	352.8	20.8	866	12	BF570994
C 13	332	19.6	434	13	BI828640
C 14	217.6	12.8	238	9	AI640509
C 15	190	11.2	803	17	A0746664
C 16	186.8	11.0	1016	11	AK007113
C 17	140.6	8.3	279	13	BI828773
C 18	135	8.0	600	10	BB615022
C 19	133.4	7.9	217	13	BI550827
C 20	116	6.8	508	9	AU126092
C 21	114	6.8	1136	13	BM459919
C 22	114.4	6.7	610	13	BI830400
C 23	112	6.6	403	14	BM853312
C 24	112	6.6	476	14	BM742560
C 25	111.4	6.6	1219	13	BM477621
C 26	110	6.5	643	10	AW163133
C 27	108	6.4	925	14	BQ421211
C 28	108	6.4	960	12	BG387948
C 29	108	6.4	1228	10	BE543630
C 30	106	6.2	950	14	BQ214006
C 31	106	6.2	1403	12	BF339614
C 32	105.4	6.2	715	12	BG773027
C 33	104.4	6.1	762	9	AI878912
C 34	103	6.1	165	14	BM847680
C 35	103	6.1	166	14	BM739099
C 36	103	6.1	358	14	BM851331
C 37	103	6.1	360	14	BM850156
C 38	103	6.1	401	14	BM773094
C 39	103	6.1	406	14	BM773061
C 40	103	6.1	413	14	BM845409
C 41	103	6.1	422	14	BM827566
C 42	103	6.1	423	13	BI668317
C 43	103	6.1	437	14	BM784562
C 44	103	6.1	438	14	BM757362
C 45	103	6.1	439	14	BM851642

ALIGNMENTS

RESULT 1
AI659243/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AI659243
tul8b06.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251379 3',
mRNA sequence.
AI659243
AI659243.1 GI:4762813
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 532)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 463.

```
FEATURES
  source
    Location/Qualifiers
      1..532
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2251379"
        /clone_lib="NCI_CGAP_Px28"
        /sex="male"
        /dev_stage="adult"
        /lab_host="DH10B"
        /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Px22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones IDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      153 a      92 c      155 g      132 t
ORIGIN
Query Match      30.6%; Score 519.4; DB 9; Length 532;
Best Local Similarity 99.6%; Pred. No. 3.8e-137;
Matches 531; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 245 TCTCTACTTCATTCCTTGGCCCTCGCAACCCATTCTCCACTCCGCGAGCCATTTTAAA 304
DB 532 TCTCTACTTCATTCCTTGGCCCTCGCAACCCATTCTCCACTCCGCGAGCCATTTTAAA 473
QY 305 AAGATGCCCTCCCTACTTATGACTCTAAATTTGCTCTCTCTCACTCTTCCCTCAGGATA 364
DB 472 AAGATGCCCTCCCTACTTATGACTCTAAATTTGCTCTCTCTCACTCTTCCCTCAGGATA 413
QY 365 TATTTCCAATTAATATACCTAAGTACTGCGCCACCTCTGCAACCCCAATGTCACATTCGA 424
DB 412 TATTTCCAATTAATATACCTAAGTACTGCGCCACCTCTGCAACCCCAATGTCACATTCGA 353
QY 425 GTCTTACTGAACACTTACTGACTGATTTCCCGAGATCTCACTCTTCTCGCTGTACCTTG 484
DB 352 GTCTTACTGAACACTTACTGACTGATTTCCCGAGATCTCACTCTTCTCGCTGTACCTTG 293
QY 485 TGC CGGAAAAGTCAGCCCTCCACCTTCTCCCTGCTTCCACTCCCAAAATACTTCTGTTGTT 544
DB 292 TGC CGGAAAAGTCAGCCCTCCACCTTCTCCCTGCTTCCACTCCCAAAATACTTCTGTTGTT 233
QY 545 TTG CAGCTCTGAGTATTACCGTGTGGCTGTTTAAATTTCTGCCTCCATCAGAAGCA 604
DB 232 TTG CAGCTCTGAGTATTACCGTGTGGCTGTTTAAATTTCTGCCTCCATCAGAAGCA 173
QY 605 GAAACTGACTCGCGAACTATTTCATCCCGCCGATAGTAGACGCTTAAAGAAAGACGCA 664
DB 172 GAAACTGACTCGCGAACTATTTCATCCCGCCGATAGTAGACGCTTAAAGAAAGACGCA 113
QY 665 AGAAGTGGTGGGAGGACTTCAGTAACATCAGTGGCAGCCTCAATTTTATCGTTTGTG 724
DB 112 AGAAGTGGGT-GGAGGACTTCAGTAACATCAGTGGCAGCCTCAATTTTATCGTTTGTG 54
QY 725 AAACGTGGATAGTAATCCCTCTATCAGTGGCTGTTTCAGGAATAAAGTGAAA 777
DB 53 AAACGTGGATAGTAATCCCTCTATCAGTGGCTGTTTCAGGAATAAAGTGAAA 1
RESULT 2
BI830189/c
LOCUS      BI830189
DEFINITION 603072855f1 NIH_MGC_119 Homo sapiens cdna clone IMAGE:5164981 5',
mRNA sequence.
ACCESSION  BI830189
VERSION    BI830189.1 GI:15941739
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 503)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-femail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11409 row: m column: 14
High quality sequence stop: 499.
FEATURES
  source
    Location/Qualifiers
      1..503
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5164981"
        /clone_lib="NIH_MGC_119"
        /tissue_type="medulla"
        /lab_host="DH10B"
        /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NOL1;
Site_2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."
BASE COUNT      121 a      125 c      117 g      140 t
ORIGIN
Query Match      28.8%; Score 489.4; DB 13; Length 503;
Best Local Similarity 99.8%; Pred. No. 1.4e-128;
Matches 501; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 698 GTGCAGCCCTCAATTTTATCGTTTGTGAAACGTGATAGTAAATCCCTCTATCAGTGGCT 757
DB 502 GTGCAGCCCTCAATTTTATCGTTTGTGAAACGTGATAGTAAATCCCTCTATCAGTGGCT 443
QY 758 GTGCAGGAATAAGTGAAGAAACAAACAGGCTAGCTTGTTCATTAATAGTGAATGAA 817
DB 442 GTTGCAGGAATAAGTGAAGAAACAAACAGGCTAGCTTGTTCATTAATAGTGAATGAA 384
QY 818 TTAATCTGATTTTCTGCTCAGTAGAAGAAAGATGTGAATCTTGTGAAAGGAGACACATTT 877
DB 383 TTAATCTGATTTTCTGCTCAGTAGAAGAAAGATGTGAATCTTGTGAAAGGAGACACATTT 324
QY 878 TTTTAAATATATGCTGTGTAAGCGGATCAGAGGAGGTCCTCCCATGGAGCACACCTCG 937
DB 323 TTTTAAATATATGCTGTGTAAGCGGATCAGAGGAGGTCCTCCCATGGAGCACACCTCG 264
QY 938 CCCTAAACATGCTGAACCCGGCTGCCATAGCCTGCTGCTGCTCCCTCCCAAGTGACTGCTC 997
DB 263 CCCTAAACATGCTGAACCCGGCTGCCATAGCCTGCTGCTGCTCCCTCCCAAGTGACTGCTC 204
QY 998 CGACAAAGGAGTACGCTCTTCAACACGATACGTTTAAAGGCAATTCAGAAACCCCTCGCT 1057
DB 203 CGACAAAGGAGTACGCTCTTCAACACGATACGTTTAAAGGCAATTCAGAAACCCCTCGCT 144
QY 1058 GTGCCGCGACTACAGCCCATTAAGAAAAAGACGACTCTATGCCCCCGCGTAAATCTTCTCA 1117
DB 143 GTGCCGCGACTACAGCCCATTAAGAAAAAGACGACTCTATGCCCCCGCGTAAATCTTCTCA 84
QY 1118 GATCAGAGGACCGTATTTCGAGCTGGAGGGAGGAGCGCTTTCTTTCACGGGGGGCTA 1177
DB 83 GATCAGAGGACCGTATTTCGAGCTGGAGGGAGGAGCGCTTTCTTTCACGGGGGGCTA 24
QY 1178 AGCGGTCTTCGAGGCCCCCTTCCA 1200
```


QY 353 CCCCTCAGGATATATTCATTAATTAATACCTAAGTGAAGTCCACCTCTGCACCCAA 412
Db 467 CACTTGAGGATATATTCATTAATTAATACCTAAGTGAAGTCCACCTCTGCACCCAA 408
QY 413 TGTACATTCGAGTCTTACTGAACACTTACTGACTGATTCCTCCGAGATCTCACCCTTCTC 472
Db 407 TGTACATTCGAGTCTTACTGAACACTTACTGACTGATTCCTCCGAGATCTCACCCTTCTC 348
QY 473 GCCTGTACCTGTGCGCGGAAGTCAAGCTCAGCCCTCCACCTTCTCCCTGCTTCCACTGCC-AAA 531
Db 347 GCCTGTACCTGTGCGCGGAAGTCAAGCTCAGCCCTCCACCTTCTCCCTGCTTCCACTGCCGAAA 288
QY 532 ATACTTCGTGTTTTGTCAGCTCTGGAGTATTTACCGTCTTGGCTGTTTTAAATTTCTGCT 591
Db 287 ATACTTCGTGTTTTGTCAGCTCTGGAGTATTTACCGTCTTGGCTGTTTTAAATTTCTGCT 228
QY 592 CCATCAGAAGCAGAACTGACTCGCGAACTATTCATTCCTCCGAGATCTCAGTACGCTT 651
Db 227 CCATCAGAAGCAGAACTGACTCGCGAACTATTCATTCCTCCGAGATCTCAGTACGCTT 168
QY 652 AAAAAAGACGGAAGTGGTGGGAGGACTTCAAGTAACTCAGTGGGAGGCTCAAT 711
Db 167 AAAAAAGACGGAAGTGGTGGGAGGACTTCAAGTAACTCAGTGGGAGGCTCAAT 109
QY 712 TTTATCGTTTGAAACCTGGATAGTATTCCTCTATCACCTGGCTGTTGCAGGAATAAA 771
Db 108 TTTATCGTTTGAAACCTGGATAGTATTCCTCTATCACCTGGCTGTTGCAGGAATAAA 49
QY 772 GTGAAAACAAAACAGCTAGCTGTTGTTCAATAAATGTGAGTTGAATT 819
Db 48 GTGAAAACAAAACAGCTAGCTGTTGTTCAATAAATGTGAGTTGAATT 1

RESULT 5
A1650982
LOCUS w956006.x1 NCI_CGAP_GC6 438 bp mRNA linear EST 16-DEC-1999
DEFINITION mRNA sequence.
ACCESSION A1650982
VERSION A1650982.1 GI:4734961
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 552 Std Error: 0.00
Seq primer: -40Up from Gibco.

FEATURES
source
1..438
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2304011"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones IDs
1257096-1258631, 1469084-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 101 a 130 c 116 g 91 t
ORIGIN
Query Match 25.1%; Score 425.4; DB 9; Length 438;
Best Local Similarity 99.8%; Pred. No. 2.6e-110;
Matches 426; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 858 TTGAAAGGAGACACATTTTTTAAATATATGCTGTGTAACGATCACAAGCAGGT 917
Db 3 TTGAAAGGAGACACATTTTTTAAATATATGCTGTGTAACGATCACAAGCAGGT 62
QY 918 CCCATGGAGCACACCTCGCCCTAAACATGCTGAACCCGGCTGCCATAGCCTTGG 977
Db 63 CCCATGGAGCACACCTCGCCCTAAACATGCTGAACCCGGCTGCCATAGCCTTGG 122
QY 978 TCCCTCAAGGTGACTGCTCCGACAAAAGGTACGCTCTTCAACGATACGTTTAAGGC 1037
Db 123 TCCCTCAAGGTGACTGCTCCGACAAAAGGTACGCTCTTCAACGATACGTTTAAGGC 182
QY 1038 AATTCCAGAAACCCCTCGGCTGTCCGCGACTACACGGCCATTAAAGAAAAGACGACTCTA 1097
Db 183 AATTCCAGAAACCCCTCGGCTGTCCGCGACTACACGGCCATTAAAGAAAAGACGACTCTA 242
QY 1098 TGCCCGCGGTAAATGTTCTCAGATCACAGGACCGTATTGAGCTGGAGGGAGGGAAGC 1157
Db 243 TGCCCGCGGTAAATGTTCTCAGATCACAGGACCGTATTGAGCTGGAGGGAGGGAAGC 302
QY 1158 CTTTCTTCACGGGGGCTAAGGGCTCTCGAGCCCTTCCATCCCGGCTCGGCCGG 1217
Db 303 CTTTCTTCACGGGGGCTAAGGGCTCTCGAGCCCTTCCATCCCGGCTCGGCCGG 362
QY 1218 GTAATCCTCCCGAGCTTCGGCGCTGCTTTTTCAGCGAGACACAACCTCTGAACGT 1277
Db 363 GTAATCCTCCCGAGCTTCGGCGCTGCTTTTTCAGCGAGACACAACCTCTGAACGT 422
QY 1278 GGGGGCC 1284
Db 423 GGGGGCC 429

RESULT 6
A1268811/c
LOCUS qo40g10.x1 NCI_CGAP_Lu5 431 bp mRNA linear EST 17-NOV-1998
DEFINITION similar to contains OFR.b3 OFR repetitive element.; mRNA sequence.
ACCESSION A1268811
VERSION A1268811.1 GI:3887978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 408.

FEATURES

Location/Qualifiers
 1. .431

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1911042"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 120 a 91 c 104 g 116 t
 ORIGIN
 Query Match 24.6%; Score 418.4; DB 9; Length 431;
 Best Local Similarity 99.5%; Pred. No. 2.6e-108;
 Matches 430; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 407 ACCAATGTCACATTCGAGTCTTACTGAATCTACTTGACTGCATTTCCCGAGATCTCACCT 466
 Db 431 ACCAATGTCACATTCGAGTCTTACTGAATCTACTTGACTGCATTTCCCGAGATCTCACCT 372
 QY 467 CTCTCCGCTGTACCTGTGCGGGAAGTCAGCCCTCCACCTCTCCCTGCTTCCACTC 526
 Db 371 CTCTCCGCTGTACCTGTGCGGGAAGTCAGCCCTCCACCTCTCCCTGCTTCCACTC 312
 QY 527 CCAAAATACCTCGTGGTGTTCGAGCTGTGGAGTATTTACCGTGTGGCTGTTTAAATTC 586
 Db 311 CCAAAATACCTCGTGGTGTTCGAGCTGTGGAGTATTTACCGTGTGGCTGTTTAAATTC 252
 QY 587 TGCCTCATCAGAGGAGGAGAACTGACTCGGAACTATTCATCCCGAGCCGATAGTA 646
 Db 251 TGCCTCATCAGAGGAGGAGAACTGACTCGGAACTATTCATCCCGAGCCGATAGTA 192
 QY 647 CGCTTAAAAAGAACGGAAGGTGGTGGAGGACTTCAGTAACATCAGGTGGCAGCC 706
 Db 191 CGCTTAAAAAGAACGGAAGGTGGTGGAGGACTTCAGTAACATCAGGTGGCAGCC 133
 QY 707 TCAATTTTATCGTTGTGAAACGTGGATAGTAATCCCTCTATCAGCTGGCTGTTGAGGA 766
 Db 132 TCAATTTTATCGTTGTGAAACGTGGATAGTAATCCCTCTATCAGCTGGCTGTTGAGGA 73
 QY 767 ATAAGTGAATAAACAACAGCTAGCTGTTTCAATAAATGAGTTGAATTAATCTG 826
 Db 72 ATAAGTGAATAAACAACAGCTAGCTGTTTCAATAAATGAGTTGAATTAATCTG 13
 QY 827 ATTTGTGCTCAG 838
 Db 12 ATTTGTGCTCAG 1

RESULT 7

AI467820/c

LOCUS

DEFINITION

tj78b11.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:2147613 3' similar to contains OFR.b3 OFR repetitive element
 ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 428)

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

RESULT 8	AI208001/C	428 bp	mRNA	linear	EST 30-NOV-1998
LOCUS	g960e06.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1839586				
DEFINITION	3', mRNA sequence.				
ACCESSION	AI208001				
VERSION	AI208001.1	GI:3769943			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 428)				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.				
FEATURES	cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 603 Std Error: 0.00 Seq primer: -400P from Gibco High quality sequence stop: 418. Location/Qualifiers 1..428 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="IMAGE:1839586" /sex="male" /lab_host="DH10B" /note="Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTCGGAGCGCGCCCAATTTTTTTTTTTT 3', Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo." 117 a 88 c 109 g 114 t				
BASE COUNT	117 a	88 c	109 g	114 t	
ORIGIN					
Query Match	23.9%;	Score 405.8;	DB 9;	Length 428;	
Best Local Similarity	98.1%;	Pred. No. le-104;			
Matches 421;	Conservative 0;	Mismatches 7;	Indels 1;	Gaps 1;	
QY	353	CCCTCAGGATATATTC	CAATTAATACCTAAGT	GACTGCCACCTCTGCAACCCAA	412
Db	428	CACCTGAGATATATTC	CAATTAATATACGTAGT	GACTGCCACCTCTGCAACCCAA	369
QY	413	TGTCACATTCGAGCT	TTACTGAACTACTT	GACTGCAATTTCCGAGATC	CACTCTTC 472
Db	368	TGTCACATTCGAGCT	TTACTGAACTACTT	GACTGCAATTTCCGAGATG	CACCTCTTC 309
QY	473	GCCTGTACCTCTG	CGGAAAGTCAGCCCT	CACGTTCTCCCTGCTTCCACTCCCAAA	532
Db	308	GCCTGTACCTCTG	CGGAAAGTCAGCCCT	CACGTTCTCCCTGCTTCCACTCCCAAA	249
QY	533	TACTTCGTGTTT	GCAGCTCTGGAGTATTT	ACCGTGTTGGCTGTTTAATTTCTGCCTC	592
Db	248	TACTTCGTGTTT	GCAGCTCTGGAGTATTT	ACCGTGTTGGCTGTTTAATTTCTGCCTC	189
QY	593	CATCAGAAGGCAGAA	ACTGACTCGCGAACTAT	TCCATCCCGCCGATAGTAGACGCTTA	652
Db	188	CATCAGAAGGCAGAA	ACTGACTCGCGAACTAT	TCCATCCCGCCGATAGTAGACGCTTA	129
QY	653	AAAAAAGAGGAAGG	TGGTGGGAGGACTT	CAGTAACATCAGGTGGCAGCTCAATT	712
Db	128	AAAAAAGAGGAAGG	TGGTGGGAGGACTT	CAGTAACATCAGGTGGCAGCTCAATT	70
QY	713	TTATCGTTTGTGAA	CGTGTAGTAGTAAT	CCCTCTATCACGTGCTGTTCAGGAATAAAG	772
Db	69	TTATCGTTTGTGAA	CGTGTAGTAGTAAT	CCCTCTATCACGTGCTGTTCAGGAATAAAG	10
QY	773	TGAAAAAAC	781		
Db	9	TGAAAAAAC	1		
RESULT 9	N67996/C	412 bp	mRNA	linear	EST 13-MAR-1996
LOCUS	yz92e04.sl Soares_multiple_sclerosis_2NBHMP Homo sapiens cDNA				
DEFINITION	clone IMAGE:290526 3', mRNA sequence.				
ACCESSION	N67996				
VERSION	N67996.1	GI:1224157			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 412) Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Willson, R. The WashU-Merck EST Project Unpublished (1995) Contact: Willson, R.K. Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: ml3 -40 forward High quality sequence stop: 355. Location/Qualifiers 1..412 /organism="Homo sapiens" /db_xref="GDB:3906366" /db_xref="taxon:9606" /clone_lib="IMAGE:290526" /sex="male" /tissue_type="multiple sclerosis lesions" /dev_stage="Age 46" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pTT3D (Pharmacia) with a modified polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI ; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTCGGAGCGCGCCCAATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTT3 vector (Pharmacia). Library went through one round of normalization to Cot5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH)." 114 a 86 c 99 g 113 t				
BASE COUNT	114 a	86 c	99 g	113 t	
ORIGIN					
Query Match	23.5%;	Score 399.4;	DB 14;	Length 412;	

AUTHORS	TITLE	JOURNAL	COMMENT
NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap .		
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
Unpublished (1997)			
Contact: Robert Strausberg, Ph.D.			
Email: cgapbs-r@mail.nih.gov			
Tissue procurement: Michael J. Brownstein, M.D., Ph.D., Michael.R. Emmert-Buck, M.D., Ph.D.			
CDNA Library Preparation: M. Bento Soares, Ph.D.			
CDNA Library Arrayed by: Greg Lennon, Ph.D.			
DNA Sequencing by: Washington University Genome Sequencing Center			
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov			
Seq primer: -40UP from Gibco			
High quality sequence stop: 342.			

```

FEATURES
source
Location/Qualifiers
1..378
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3309359"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
100 a 78 c 100 g 100 t
BASE COUNT
ORIGIN
Query Match
Best Local Similarity 20.9%; Score 354.2; DB 12; Length 378;
Matches 370; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

```

BASE COUNT	100 a	78 c	100 g	20 9;	Score 354.2;	DB 12;	Length 378;
ORIGIN				Best Local Similarity	97.6%;	Pred. No. 5.6e-90;	
				Matches 370;	Conservative 0;	Mismatches 8;	Indels 1; Gaps 1;
Query Match							
404	GCAACCCAAATGCACATTCGAGTCTTACTGAACTACTTGACTGCATTTCCCGAGATCTCA	463					
dbb	378	GCAACCCAAATGCACATTCGAGTCTTACTGAACTACTTGACTGCATTTCCCGAGATCTCA	319				
464	CTCTTCTCGCCTGTACCCCTGTGCGCGGAAAGTCAGCCCTCCACCTTCCTCCCTGCTTCCA	523					
dbb	318	CTCTTCTCGCCTGTACCCCTGTGCGCGGAAAGTCAGCCCTCCACCTTCCTCCCTGCTTCCA	259				
424	CTCCCAAAATACTTCCTGGTGTGAGCTCTCGAGTATTTACCGTCTTGCGTCTGTTTAAAT	583					
dbb	258	CTCCCAAAATACTTCGTGTGTTTGCAGCTATGGAGTATTTACCGTCTTGCGTCTGTTAAAT	199				
584	TTCGCTCCCATCAGAAAGCAGAACTGACTCGGAACTATTCATCCCGACCGCATAGT	643					
dbb	198	TTCGCTCCCATCAGAAAGCAGAACTGACTCGGAACTATTCATCCCGACCGCATAGT	139				
644	AGACGCTTAAAAAACAACGGAAGGTGGGTGGAGGACTTCAGTAACATCAGGTGGCA	703					
dbb	138	AGACGCTTAAAAAACAACGAAGAGGTGGGT- GGAGGACTTCAGTAACATCAGGTGGCA	80				
704	GCCTCAATTTTATCGTTGTGTAACCGTGGATAGTAAATCCCTCTATCACGTGGCTGTTGCA	763					
dbb	79	GCCTCAATTTTATCGTTGTGTAACCGTGGATAGTAAATCCCTCTATCACGTGGCTGTTGCA	20				
764	GGAATAAAGTGA AAAACA	782					
dbb	19	GGAATAA AAAAAGCCCA	1				

LOCUS	BF570994	866 bp	mRNA	linear	EST 12-DEC-2000
RESULT 12					
BF570994/c					

602075915F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4243371 5',
mRNA sequence.
BF570994
BF570994.1 GI:11644706
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 866)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: ATCC/DCTD/DTP
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM1053 row: m column: 04
High quality sequence stop: 645.
Location/Qualifiers
1 . 866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4243371"
/clone_lib="NIH_MGC_62"
/tissue_type="melanotic melanoma, high MDR"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccctcgccc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGCCGATG-dfr(30)BN-3'
(where B = A, C, G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
240 a 177 c 259 g 189 t 1 others

	Query Match	20.8%;	Score 352.8;	DB 12;	Length 866;
	Best Local Similarity	85.8%;	Pred. No. 1.8e-89;		
	Matches 557;	Conservative 0;	Mismatches 68;	Indels 24;	Gaps 1
701	GCAGCCTCAATTTTATCGTTTGTGAAACGTGGATAGTAAATCCC-TCTATCACTGCGTGT 759				
Db	638	GCACCTCAATTTTATCGTTTGTGAAACTGGATTAGTAATCCCTTCTATCACTCGCTGT 579			
760	TGCAGGAATAAGTGTAAAAACAACACAGGCTAGCTTCTTCAATAAAT-GTGAGTGTGAAT 818				
Db	578	TTACAGGAATAAGTGTAAACAACACAC--GGCTAGCTTGTCCATAAATGGTGAGTTGAAT 522			
819	TAAATCTCATTTGGTGCAGTAGAAAGAATGTGAATACTTTGGRAAGA---AGCACAT 875				
Db	521	TAAATCTCATTTGGTGTCA--TAGACAAGATTGTGAATACTTTGGAAAGGAAGACACAGC 464			
876	TTTTTTAAATATATGCCCTGGTAAACCGGATCAGAAGGC-AGGTCCCATGGAGCACACCC 934				
Db	463	TTTGTGAATATATGCCCTGGTTAAACCGATCAGAAGGCAGGTCCCATGGAGCACACCC 404			
935	TCGGCCTTAAACATGCTGAACCGGGCTGCCATAGCCTCGCTGGTCCCTCCCAAGG----T 989				
Db	403	TCGGCCT--AACAATGCTGAACCGGGCTTGCCATACCTCGCTGGTCCCTCCACAGGTGACT 345			
990	GACTGCTCCGACAAAGGGTAGCTCTTCAACCGATACGTTTAAAGCAATTCAGAAAC 1049				
Db	344	TGCTCCGGTACAAAGGGTAGCTCTTTCACCGCATACGTTTAAAGGCCATTCACGTAAC 285			

